

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:47:19 ; Search time 3552.46 Seconds
(without alignments)
5295.174 Million cell updates/sec

Title: US-09-985-689A-2

Perfect score: 2245

Sequence: 1 NDVARGIVKADVAQSSGYLY.....EVOAYNVFVGQNFSLAIVN 434

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0985689/runat_31032004_161806_4155/app_query.fasta_1.3498
-DS=genEmbl -QMT=fastap -SUFFIX=fgc -MINMATCH=0.1 -DOOPCL=0 -DOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-CUTFM=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0985689@cgn 1.11925 @runat 31032004 161806 4155 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LORGLQ
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPEXT=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rcd.*
36: em_htg_nam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2237	99.6	1920	1	AB046403	Bacillus
2	2237	99.6	1920	6	AB368116	Sequence
3	2197	97.9	1823	6	AB368118	Sequence
4	2196	97.8	1823	1	AB084155	Bacillus
5	2191	97.6	1305	6	AX839476	Sequence
6	2191	97.6	1823	1	AB051423	Bacillus
7	2191	97.6	1823	6	AB368117	Sequence
8	2130	94.9	1302	1	AB046406	Bacillus
9	2118.5	94.4	3003	6	AR069954	Sequence
10	2118.5	94.4	3003	6	BD062155	Nucleic a
11	2010.5	89.6	1299	1	AB046405	Bacillus
12	2008.5	89.4	1299	1	AB046402	Bacillus
13	1999.5	89.1	1299	1	AB046404	Bacillus
14	1998.5	89.0	2218	6	E03808	EO3808 DNA encodin
15	626.5	27.9	60006	1	AF268611	Unculture
16	571	25.4	301305	1	AE017218	Geobacter
17	526.5	23.5	5890	3	DDU60086	Dictyosteli
18	509.5	22.7	5772	3	AF466309	Dictyosteli
19	490	21.8	6115	3	DDU20432	Dictyosteli
20	457.5	20.4	1977	6	AR201152	Sequence
21	443	19.7	5109	3	AF263455	Dictyosteli
22	428.5	19.1	126928	2	AC096673	Trypanoso
23	425.5	19.0	1236	6	AR201146	Sequence
24	425.5	19.0	1962	6	AR201155	Sequence
25	425.5	19.0	12452	1	AE010265	Pyrococcu
26	418	18.6	14376	1	AE013049	Thermoana
27	407.5	18.2	303450	1	SC0939130	Streptomy
28	387	17.2	299850	1	AP004601	Oceanobac
29	365	16.3	3348	1	D83672	Streptomyce
30	360.5	16.1	300425	1	AF005044	Streptomy
31	359.5	16.0	2539	6	AR202322	Sequence
32	357.5	15.9	292550	1	AP001513	Bacillus
33	354.5	15.8	2809	6	AR202321	Sequence
34	354.5	15.8	3900	1	AB007809	Streptomy
35	351	15.6	291000	1	SC0939105	Streptomy
36	345.5	15.4	302300	1	AP005034	Streptomy
37	343	15.3	135638	1	AF484556	Streptomy
38	340	15.1	1239	1	AF305633	Thermoana
39	340	15.1	1239	1	AY028704	Thermoana
40	339	15.1	12395	1	AE013026	Thermoana
41	338.5	15.1	1329	6	AX433519	Sequence
42	331.5	14.8	301150	1	AP004602	Oceanobac
43	331	14.7	207829	1	BSUB0010	Bacillus su
44	329.5	14.7	300800	1	SC0939112	Streptomy
45	323.5	14.4	2731	1	AF015225	Thermococ

ALIGNMENTS

RESULT 1

Db 1816 AATGATTTATTAATTCGCCCAAGTGCACATATACCATTTGAGTGGCAGCATATAAT 1875
Qy 421 ValProValGlyProGlnAsnPheserLeuAlaIleValAsn 434
Db 1876 GTGCGGTTGGACACCAAACTTCTCGTTGGCAATTTGTGAAC 1917
RESULT 2
AR368116
LOCUS AR368116 1920 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 3 from patent US 6376227.
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1920)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S. and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 3 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1920
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores: 1,05e-135 Length: 1920
Pred. No.: 2327.00 Matches: 433
Score: 99.77% Conservatives: 0
Percent Similarity: 99.77% Mismatches: 1
Best Local Similarity: 99.74% Indels: 0
Query Match: 99.64% Gaps: 0
DB: 6
US-09-985-689A-2 (1-434) x AR368116 (1-1920)
Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrClyLeuTyr 20
Db 616 AATGATGTGCCAGAGATTTGTCAAGCGGATGTGGCAGACAGAGCTACCGTTTGTAT 675
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGCACAGCCAGATTGTCCAGTGTCCGATATCGATTGGATACAGGAAGAACACAGT 735
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATCAAGCCTTCGCGGTAAATAACAGCACTATATGCACTGGGTGCGACGAAT 795
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCCAATGATACGACAGCGTATGTACCATGTGGCAGGTTCCGTTATAGGAATGGC 855
Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAATAAGAAATGGCACTCAAGCGAATCTGGTTTCAATCCATCATGTGATAGC 915
Qy 101 SerGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 916 AGTGGTGGGCTTGAGGCTTCCTTCAATCTCAAACTTATTCAGCAAGCAATTCAGT 975
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThrThr 140
Db 976 GCAGGTGCCAGAAATTCATACAACTCTGGGGGCGAGCGGTGAATGGGCTACAGACA 1035
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1036 GAITTCAGAAATGTGGATGACTATGAAGGAAAAATCATATGACGATTTCTTTCGCGGT 1095
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1096 GGGNATGAAAGCCGACCGCGGTACCATCAGTGCACCTGGTACCGCTAAACACGCCATA 1155

Qy 181 ThrValGlyValAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTCGGGCGCAACCGAAACCTGCGTCCAAAGCTTCGGTTCCCTATGCAGATAATATAAC 1215
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CACGTTGCACAGTTCTTCCCGTGGCCGCAAAAGATGGCGGAATCAAGGCTGATGTC 1275
Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTTCTTTCGCCCGATTCCTCCTTC 1335
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGCAATCATGACAGCAATATGCCTATATGGGTGGAACTCCATGTCACACCGAT 1395
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1396 GTTGGCGGGAATGTTCACAGCTCCCGTAGCATTTTGTGAAAAATAGAGGATCACTCCT 1455
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 1456 AAGCCTTCCCTATTGAAGCAGCTTTGATTGCGAGTGTCTGCTGATGTTGGATTGGTTAT 1515
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAACGGAACCAAGATGGGCGGAGTGACCTCGATAAATCGTTGAACGTTGCTAT 1575
Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 1576 GTGAACGAATCCAGTGCCTTATCAACTAGCCAAAAGCGACATATACCTTTACTGCAACG 1635
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1636 GCGGGCAAGCCATTGAAATCTCCCTGGTATGTGCGATGCCCTCGAAGCCTACTGCT 1695
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db 1696 TCTGTAAACCTGGTCAATGATTGGATTGGTCAATACAGCAACCAACGAAACAGATAT 1755
Qy 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400
Db 1756 GTCGGGAAATGACTTCTCAGCACCATTTGACAATAACTGGGATGGCCGCAATAACGAGAA 1815
Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGATTTATTAATTCGCCCAAGTGAACATATACCATTCAGGTGCAAGCATATAAT 1875
Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1876 GTGCGGTTGGACCAAACTTCTCGTTGGCAATTTGTGAAC 1917

RESULT 3
AR368118 1923 bp DNA linear PAT 12-SEP-2003
LOCUS AR368118
DEFINITION Sequence 7 from patent US 6376227.
ACCESSION AR368118
VERSION AR368118.1 GI:34601779
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1923)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S. and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 7 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1923
/organism="unknown"
/mol_type="genomic DNA"

Alignment Scores:		4.73e-133	1923
Pred. No.:	Score:	2196.00	420
Percent Similarity:	Conservative:	99.54%	12
Best Local Similarity:	Mismatches:	96.77%	2
Query Match:	Indels:	97.82%	0
DB:	Gaps:	1	0
US-09-985-689A-2 (1-434) x AB084155 (1-1923)			
QY	1	AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyGlyLeuTyr	20
DB	619	AATGATGTTGCAGTGGAAATTGTCAGGCGATGTCGCGCAGAGCAGTACGGGTTGTAT	678
QY	21	GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40
DB	679	GGCAAGGACAGATCGTAGCGGTTCGCGATACAGGCGTTGATACAGGTGCGCAATGACGT	738
QY	41	SerMetHisGluAlaPheAspGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60
DB	739	TCGATGCGATGAGCTTCGCGGGGAAATTAATCTGCTATATGATTTGGACGCAAT	798
QY	61	AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly	80
DB	799	AATGCCAATGATACCAATGCTCATGTGTCAGCATGTGCTGCTCCGTTATTAGGAAACGGC	858
QY	81	AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	100
DB	859	TCCACTAATTAAGAAATGGCGCTCAGCGCAATCTAGTCTTCCAAATCTATCATGGATAGC	918
QY	101	SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer	120
DB	919	GGTGGGGGACTTGGAGGACTACCTTCGAATCTGCAAACTTATTTCAGCAAGCATACAGT	978
QY	121	AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValLeuGlyAlaTyrThrThr	140
DB	979	GCTGGTGCAGAAATTCATACAAATCTCTGGGGAGCAGCAGTGAATGGGGTTACACAACA	1038
QY	141	AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla	160
DB	1039	CATTCCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCCTTTTCGTGCC	1098
QY	161	GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	180
DB	1099	GGGAATGAAGACCGCAACCGCGCAACCATCATGTCACCGAGCAGCAGTAAATAATGCAATA	1158
QY	181	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn	200
DB	1159	ACAGTCGGAGTACGGAAACCTCGCCCAAGCTTCGGTCTTATGCGACAAATATCAAC	1218
QY	201	HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal	220
DB	1219	CATGTGGCACAGTTCTCTTTCACGTGGACCGCAAGGATGGACGGATCAACCGGATGTC	1278
QY	221	MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe	240
DB	1279	ATGGCACCGGACGCTTCATCTACTATCAGCAGATCTTCTTTGACCGGATTCCTCTTC	1338
QY	241	TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle	260
DB	1339	TGGGGAACCATGACAGTAAATATGCATACATGCTGGTGAACGTCCTCGCTTACACCGATC	1398
QY	261	ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro	280
DB	1399	GTGTGTGGAACGTGGCAGCTTCGTGAGCATTTTGTGAAACACAGAGGATCACACCA	1458
QY	281	LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr	300
DB	1459	AAGCCTTCTCTATTAAAGCGGCACTGATTGCGGTGCGAGCAGCATCGGCTTGGCTAC	1518
QY	301	ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr	320
DB	1519	CCGAACGGTAACCAAGGATGGGACAGTGCATTTGGATAATCCCTAAACGTTGCCTAT	1578
QY	321	ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr	340
DB	1579	GTGAACGAGTCCAGTTCTTATCCACAGCCAAAGCGAGTACTCTGTTTACTGCTACT	1638
QY	341	AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla	360
DB	1639	GCCGGCAAGCCTTTCAGAGATCTCCTGTGTATGCTGATGCCCTGCGAGCAACTGCT	1698
QY	361	SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrArgTyr	380
DB	1699	TCCGTAAACGCTTGTCAATGACCTGCGCTTGTCAATACCGCTCCAAATGGCACAATAT	1758
QY	381	ValGlyAsnAspPheSerAlaProPheAspAsnTrpAspGlyArgAsnValGlu	400
DB	1759	GTTGGAATGACTTACTTCGCCATACATAATACTAGGATGCCCGCAATAACGTAGAA	1818
QY	401	AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn	420
DB	1819	AATGTATTTATTAATGCGCCCAAGCGGACGTATACAAATTGAGGTACAGCTTATAAC	1878
QY	421	ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn	434
DB	1879	GTACCGGTTGACCCACAGACCTTCTCGTTGCAATTGTGAAC	1920
RESULT 5			
LOCUS	AX839476	1305 bp	DNA linear PAT 15-DEC-2003
DEFINITION	Sequence 2 from Patent EP1347044.		
ACCESSION	AX839476		
VERSION	AX839476.1	GI:39922766	
KEYWORDS	Bacillus sp. KSM-KP43		
SOURCE	Bacillus sp. KSM-KP43		
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
REFERENCE	1		
AUTHORS	Okuda, M.K., Saito, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K., Saeiki, K.K., Kobayashi, T.K. and Nomura, M.K.		
TITLE	Alkaline protease		
JOURNAL	Patent: EP 1347044-A 2 24-SEP-2003;		
FEATURES	Kao Corporation (JP)		
source	Location/Qualifiers		
CDS	1..1305		
	/organism="Bacillus sp. KSM-KP43"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:109322"		
	<1..1305		
	/note="unnamed protein product"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAE85146.1"		
	/db_xref="GI:39922767"		
	/translation="NDVARGIVKADVAVQSSVGLYGGQIVAVADTGLDGRNDSMHE		
	AFRGKITAIYALGRNTNANDNGHGHVAGSVLGNSTNGKMAPOANLVFQSIMDSGG		
	GIGGLPSNLQTLFSQAYAGARITHTNSWGAAVNGAYTTDSRNVDDYVRKNDMTILFAA		
	GNENPGNGTISAPGTAKVATVGTATENLRPSFYADINHVQSPSSGPTKDGKIKP		
	DYVAPGFTLSRSSLSAPTSFWANHDSKYAMGVTSMATPTVAGNVQALREHFVYKR		
	GITPKPSLLKALIAAGADIGLYGPNQGWGRVTLDKSLNVAVYNESSLSISQAT		
	YSEFTACKPLKISLVNSDAPASTTASVTLVNDLVLITAPNGTQYVGNDFTSFYNDN		
	WDGRNVENVFINAFQSGYITIEVQAVNVPVGPQFSLAIVN"		
ORIGIN			
Alignment Scores:	6.17e-133	Length:	1305
Pred. No.:	2191.00	Matches:	419
Score:	99.54%	Conservative:	13
Percent Similarity:	96.54%	Mismatches:	2
Best Local Similarity:	97.59%	Indels:	0
Query Match:	6	Gaps:	0
US-09-985-689A-2 (1-434) x AX839476 (1-1305)			
QY	1	AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyGlyLeuTyr	20

1 AATGATGTTGGCGTGAATTTCTCAAGCGGATGTGGCTCAGACGAGCTACGGGTTCTAT 60
 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyValGlnAspSer 40
 61 GGACAGGACAGATCGTAGCGTTCGCCATACAGGCTTGATACAGGTGCGCAATGACAGT 120
 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyValThrAsn 60
 121 TCGATGATGAAGCCTTCGCGGGGAAATTTACTGCATTATATGCAATGATGCGGACGCGAAT 180
 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 181 AATGCCAATGATACCAATGCTCATGCTAGCGATGCTGGCTCGCTCGATTAAGAAACGCG 240
 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 241 TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAACTATCATGATGATAGC 300
 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 301 GTGGGGGACTTGGAGGACTCTTGGAACTCTGCAACCTTATTGAGCAAGCATACAGT 360
 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValAlaThrThr 140
 361 GTGTGCGCCAGAACTCATACAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 420
 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 421 GATTCAGAAATGCGATGACTATGTGCGCAAAATGATGATGACCATCTTTTCGCTGCC 480
 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 481 GGAATGAGGACCGAACCGCGGACCACTCAGTGCCACAGGACAGCAGTAAATGCAATA 540
 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 541 ACAGTCGGAGTACCGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGACAAATATCAAC 600
 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyValArgIleLysProAspVal 220
 601 CATGTGGCAGATTCCTTCATGTCGACCGACCAAGATGGAGCGATCAACCGGATGTC 560
 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 661 ATGGACCGGGAACGTTCTACTATCATCAGCAAGATCTCTCTGACCGGATTCCTCTTC 720
 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 721 TGGGCGGAACCATGACAGTAATATGCATATCATGGTGGACGTCCTCATGCTACCGATC 780
 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 781 GTTCCTGGAAACGTTGGCACAGCTTCGTGAGCAATTTGTGAAAAACAGAGGCAATCACCA 840
 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
 841 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGTGCAGTCACTGACATGACGCTTGCTAC 900
 301 ProAsnGlyAsnGlnGlyTyrGlyValThrLeuAspLysSerLeuAsnValAlaTyr 320
 901 CCGAACGGTAACCAAGGATGGGACGAGTGACATTTGATTAATCCCTGAACTGCTGCTAT 960
 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
 961 GTGAACGAGTCCAGTCTCTATCCACCGCAAAAGCGAGTACTCGTTTACTGCTACT 1020
 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 1021 CCGGCAAGCCTTTGAAATCTCCTGCTGTAAGTCTGATGCTGCGGACCAACTGCT 1080
 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrArgTyr 380

1081 TCCTGAACGCTTGTCAATGATCTGGACCTTGTCTATTCCGCTCCAAATGGCACAGTAT 1140
 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
 1141 GTAGGAATGACTTTACTTTCGCATACATGATACTGGATGGCGGCAATTAACGTAGAA 1200
 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 1201 AATGATTTATTAAATGCACCAACAGCGGACGTATACAAATGAGGTACAGCTTATAAC 1260
 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 1261 GTACCGTGTGGACACAGACCTTCTCGTTGGCAATGTGAAT 1302

RESULT 6
 AB051423
 LOCUS
 DEFINITION
 AB051423
 ACCESSION
 VERSION
 AB051423.2 GI:20521154
 KEYWORDS
 SOURCE
 ORGANISM
 Bacillus sp. KSM-KP43
 Bacillus sp. KSM-KP43
 Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.

REFERENCE
 1 Itoh, S. and Saeki, K.
 new protease
 TITLE
 JOURNAL
 2 (bases 1 to 1923)
 AUTHORS
 Saeki, K.
 DIRECT SUBMISSION
 Submitted (21-NOV-2000) Katsuhisa Saeki, KAO CORPORATION;
 2606, AKABANE, ICHIKAINAKHI, HAGA, TOCHIGI 321-3486, Japan
 (E-mail: 13871850ka@kao.co.jp, Tel: 81285687471 (ex. 7471),
 Fax: 81285687403)
 On May 9, 2002 this sequence version replaced gi:14164344.

COMMENT
 FEATURES
 source
 Location/Qualifiers
 1..1923
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="genomic DNA"
 /strain="KP43"
 /db_xref="taxon:109322"
 1..1923
 /gene="PROF"
 1..1923
 /gene="PROF"
 /codon_start=1
 /transl_table=1
 /product="protease"
 /protein_id="BAB55674.2"
 /db_xref="GI:20521155"
 /translation="MRKKKVFSLVLSAAAILSTVALSNPSAGARNFDLDFKGIQTT
 TDAGFSKQGTGAALFVSENVKLPKGLKLETPANNKLHIQFNGPILBETKQ
 QLEKTAKILDYIFDYAYIVEYEDVKSATSTIEHVSVPYLPYRIDPQFTKAS
 ELKAVALDTRKQKNEVQLRGEIQIAFALSNVLYITAKPEYKVMNDVARGIVKADV
 AQSSGYLGGQIVAVADTGLDTRGNSDWEAPRGKITALIALGRTNANDTNGHGT
 HVAQSVLNGSTNGKMAQANLVQSNDSGGGLGGLPSNLQTLFQSAAGAKIHTN
 SWGAANGAVYTTDSRNDVYVRKNDMTILFAAGNEGNGGTISAPGTAKNAITVGAH
 NLRPSRYADNINHVAFSSRGPTKGRIPKPDVMAQPTTILSARSLAPDSSFWAH
 DSKAYNGGTSMATPIVAGNVAQLREHFVNRGJTPKPSLLKALILAGADIGLYEN
 GQGWNRVILDSLNVAKYNESLSSTSQKATVSTATAGKPKILSLWSDAPASTTA
 SVTLNLDLVLITAPNGQYVGVNDFTSFYNDNDGRNVNFVFNAPQSCTYITIEVQA
 YNVPVGPQTFSLAIVN"

ORIGIN

Alignment Scores:
 Pred. NO.: 9,950-133 Length: 1923
 Score: 2191.00 Matches: 419
 Percent Similarity: 99.54% Conservative: 13
 Best Local Similarity: 96.54% Mismatches: 2
 Query Match: 97.59% Indels: 0
 DB: 1 Gaps: 0

```
US-09-985-689A-2 (1-434) x AB051423 (1-1923)

QY 1 AsnAspValAlaAargGlyLeuValValAlaAspValAlaGlnSerSerTyGlyLeuTyr 20
Db 619 AATGATGTTGGCGCTGGAATTTCTCAAGCGGATGTGGCTCAGACAGCTACGGGTTGTAT 678
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAGACAGATCGTAGCGGTTGCCGATACAGGCTTTGATACAGTTCGCAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLeuValAlaLeuValAlaLeuValAlaLeuValAla 60
Db 739 TCGATGCAATGAAGCTTCCCGGGGAAAAATTAAGTCAATATATGCAATTTGGGACGACGAT 798
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACAAATGGTCTGATGACGATGCTGGCTGCTGCTGCTGCTGCTGCTGCTG 858
QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAATCTATCATGATGATG 918
QY 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 919 GGTGGGGACATTGGAGGACTACCTTCGAATCTGCAACCTTATTTCAGCCAAGCATACAGT 978
QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
Db 979 GCTGCTGCAGATTCATCAAACTCTCGGGAGCAGCAGTGAATGGGCTTACACACACA 1038
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1039 GATTCAGAAATGTGGTACGATATGTGGCGAAAAATGATATGACGATCCCTTTTCGCTGCC 1098
QY 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GGAATGAGGACCGGACCGGGAACCATAGTGCACCGCAGCAGCTAAATATGCATATA 1158
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGCTACGGAAAACTCGCCCAAGCTTTGGGTCTTATGGGCAATATCAAC 1218
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCACAGTTCTCTTCAGTGGCCGACCAAGAGGATGGACCGATCAACCCGGATGTC 1278
QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGGCACCGGAACTTCATATCATCAAGATCTTCTTGCACCGGATTCCTCCTTC 1338
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1339 TGGGCAACCATGACAGTAATATCATCATGATGGGTGGAACGTCATGGCTTACACCGATC 1398
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1399 GTTGTGGAAAGTGGCAGCTTCGTGAGCATTTTGTGAAAACAGAGGATCACACCA 1458
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 1459 AAGCCCTCTCTATTAAGCGGCACTGATTCGGGTGACGTGACGTGCGCTTGGGTAC 1518
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGCATTTGGATTAATCCCTGAAAGTTGGCTAT 1578
QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 1579 GTGACGAGTCCAGTCTCTATCCACCAGCCAAAAAGCGAGTACTCTGTTACTTGTACT 1638
QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThra 360
Db 1639 GCCCGCAGCCCTTGAATAATCTCCCTGGTATGCTGATGCCCTGCCGAGCACAACTGCT 1698
```

```
QY 361 SerValThrLeuValAlaAsnAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db 1699 TCCGTAAGCGCTGTCAATGATCTGACCTTGTCAATCCGCTCCAAATGGCACAGAT 1758
QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTyrAspGlyArgAsnValGlu 400
Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATACTGGGATGGCGCAATAACGTAGAA 1818
QY 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1819 AATGTATTATTATTAATGACCAACCAACCGGACGTATCAATTCAGGTACAGCTTATAAC 1878
QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGGTTGGACCAACAGACCTTCTCGTTGGCAATTTGGAAT 1920

RESULT 7
LOCUS AR368117 1923 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 5 from patent US 6376227.
ACCESSION AR368117
VERSION AR368117.1 GI:34601778
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1923)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S. and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 5 23-APR-2002;
FEATURES
Location/Qualifiers
source 1..1923
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 9,95e-133 Length: 1923
Score: 2191.00 Matches: 419
Percent Similarity: 99.54% Conservative: 13
Best Local Similarity: 96.54% Mismatches: 2
Query Match: 97.59% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-2 (1-434) x AR368117 (1-1923)
QY 1 AsnAspValAlaAargGlyLeuValValAlaAspValAlaGlnSerSerTyGlyLeuTyr 20
Db 619 AATGATGTTGGCGCTGGAATTTCTCAAGCGGATGTGGCTCAGACAGCTACGGGTTGTAT 678
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAGACAGATCGTAGCGGTTGCCGATACAGGCTTTGATACAGTTCGCAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLeuValAlaLeuValAlaLeuValAlaLeuValAla 60
Db 739 TCGATGCAATGAAGCTTCCCGGGGAAAAATTAAGTCAATATATGCAATTTGGGACGACGAT 798
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACAAATGGTCTGATGACGATGCTGGCTGCTGCTGCTGCTGCTGCTGCTG 858
QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAATCTATCATGATGATG 918
QY 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 919 GGTGGGGACATTGGAGGACTACCTTCGAATCTGCAACCTTATTTCAGCCAAGCATACAGT 978
QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
Db 979 GCTGCTGCAGATTCATCAAACTCTCGGGAGCAGCAGTGAATGGGCTTACACACACA 1038
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1039 GATTCAGAAATGTGGTACGATATGTGGCGAAAAATGATATGACGATCCCTTTTCGCTGCC 1098
QY 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GGAATGAGGACCGGACCGGGAACCATAGTGCACCGCAGCAGCTAAATATGCATATA 1158
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGCTACGGAAAACTCGCCCAAGCTTTGGGTCTTATGGGCAATATCAAC 1218
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCACAGTTCTCTTCAGTGGCCGACCAAGAGGATGGACCGATCAACCCGGATGTC 1278
QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGGCACCGGAACTTCATATCATCAAGATCTTCTTGCACCGGATTCCTCCTTC 1338
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1339 TGGGCAACCATGACAGTAATATCATCATGATGGGTGGAACGTCATGGCTTACACCGATC 1398
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1399 GTTGTGGAAAGTGGCAGCTTCGTGAGCATTTTGTGAAAACAGAGGATCACACCA 1458
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 1459 AAGCCCTCTCTATTAAGCGGCACTGATTCGGGTGACGTGACGTGCGCTTGGGTAC 1518
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGCATTTGGATTAATCCCTGAAAGTTGGCTAT 1578
QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 1579 GTGACGAGTCCAGTCTCTATCCACCAGCCAAAAAGCGAGTACTCTGTTACTTGTACT 1638
QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThra 360
Db 1639 GCCCGCAGCCCTTGAATAATCTCCCTGGTATGCTGATGCCCTGCCGAGCACAACTGCT 1698
```

```

Db 979 GCTGTCGTCAGAAATTCATAAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACACA 1038
Qy 141 AspSerArgAsnValAspAspTyrValArgLyAsnAspMetThrIleLeuPheAla 160
Db 1039 GATTCCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLyAsnAla 180
Db 1099 GGGAAATGAAGACCGCAACCGCGGCAACCATCAGTCACCGACGACGACGCAATGAATA 1158
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACATGCGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACATATCAAC 1218
Qy 201 HisValAlaGlnPheSerArgGlyProThrIleAspGlyArgIleIleysProAspVal 220
Db 1219 CATGTGGCACAGTTCTCTTACGTGGACCGACCAAGGATGGAGGATCAACCGGATGTC 1278
Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGGCACCGGAGCTTCATATCATATCAGCAAGATCTCTCTCCACCGGATTCCTCCTTC 1338
Qy 241 TrpAlaAsnHisAspSerIleTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1339 TGGCGCAACCATGACAGTAATATGCAATATGCGTGGACGTCCTGCTACACCGATC 1398
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIleAsnArgGlyIleThrPro 280
Db 1399 GTTGTGAAACGTGGCACAGCTTCGTGACGATTTTGTGAAACAGAGGATCATCAACA 1458
Qy 281 LysProSerIleLeuIleValAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 1459 AAGCCTCTCTATTAAAGCGCACTGATTCGCGGTCAGCTGACATCGCGCTTGGCTAC 1518
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrIleAspLysSerLeuAsnValAlaTyr 320
Db 1519 CCGAACCGTAACCAAGATGGGACGAGTGACATTCGATAAATCCCTGAACGTGCTAT 1578
Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnIleAlaThrTyrThrPheThrAlaThr 340
Db 1579 GTGAACGATCTGCTCTATCCACCGCAAAAGCGACGACTCTGTTACTGCTACT 1638
Qy 341 AlaGlyIysProLeuIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1639 GCGCGAAGCCTTTGAAATCTCCCTGCTGATGCTGATGCTGCTGCGACCACTGCT 1698
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db 1699 TCCGTAAACGCTGTCAATGATCTGGACCTTGTCTATACCGCTCCAAATGGCACAGTAT 1758
Qy 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400
Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATACTGGATGGCGCAATAACGTAGAA 1818
Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1819 AATGTATTATTATGATCACCACAAAGCGGACGATACAAATGAGGTACAGGCTTATAAC 1878
Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGGTTGGACACAGACTTCTCGTTGGCAATGTGAAT 1920

```

RESULT 8

```

AB046406
LOCUS AB046406 1302 bp DNA linear BCT 23-JAN-2001
DEFINITION Bacillus sp. NV1 PROE gene for protease, partial cds.
ACCESSION AB046406
VERSION AB046406.1 GI:12381944
KEYWORDS Bacillus sp. NV1
SOURCE Bacillus sp. NV1
ORGANISM Bacillus; Firmicutes; Bacillales; Bacillaceae; Bacillus.

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

```

```

1 (sites)
Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
Horikoshi,K.
Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
20568675
11118284
2 (bases 1 to 1302)
Saeki,K.
Direct Submission
Submitted (20-JUL-2000) Katehisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
Location/Qualifiers
1..1302
/organism="Bacillus sp. NV1"
/mol type="genomic DNA"
/strain="NV1"
/db_xref="taxon:133781"
1..1302
/genes="PROE"
<1..>1302
/genes="PROE"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21269.1"
/db_xref="GI:12381945"

```

```

gene
CDS

```

```

translation="NDVARGIVKDVAAQSSVGLYGGQVAVADTGLDTRNDSMHE
AFRGKTAIYALGRNNDPNHGTHVAGSVLNGTSKMGKPAQNLVFGSVMDNSG
GCGSPNSVTLLFSQAYAGARIHTNSWGPVNGAYTDSRVDDYVRKNDMVLFAA
GNEGNGGTISAPFAKXNAITVGTENLRFSGYADNINVAOFSSRGPTKGRIFK
DVMAPTFILSARSLPADPSFWANHDSKYAMGTSNATIVAGNVQAQLREHFKR
GTPKPSLLKAAALIAATDGLGYPGNGQWGRVTLDKSLNVPFNWTSLSLTOKAT
YFTAQSGKPLKISLVSDAPASTASVTLVNDLDELVTAPNGKYGVNDFTAPYDNN
WGRNNVENVFVAPQSGTYTVEVQAVNPQGPQAFSLAIVN"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 5,41e-129 Length: 1302
Score: 2130.00 Matches: 401
Percent Similarity: 97.93% Conservative: 24
Best Local Similarity: 92.40% Mismatches: 9
Query Match: 94.88% Indels: 0
DB: 1 Gaps: 0

```

US-09-985-689A-2 (1-434) x AB046406 (1-1302)

```

Qy 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AACGATGTTGCCAGAGGCATTTGAAAAGCCGATGTTGCCAGAGCAGCTATGTTTATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGGCAAGGGCAAGTGGTTCAGTAGCGGATACCGGACTGGATACAGGCGGTAATGACAGC 120
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGCGATGAAGCGTTCCGAGAAAGATTACCGCGATATATATGCTTTCGAAAGAACAAAC 180
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AACGCCAAATGATCCAAACGACGACGCGGATGTTGCCGATCTGTTTGGAAACCGGT 240
Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 241 ACTTCGAATAAAGGATGCTCCCAAGCTAACTAGTTTCCCAATCTGTTATGACAGC 300
Qy 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120

```

301 AATGCGGACTTGGCGGACTGCTTCCAAATGTAAGTACATTATTACGCCGACATATAGT 360
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValaTyThr 140
Db 361 CCCGGTGCAGATCCATCAGACTCATCGGAGCGCTGTAAAGGAGCGTACACTACT 420
Qy 141 AspSerArgAsnValAspAspTyrValArgIlyAsnAspMetThrIleLeuPheAlaAla 160
Db 421 GATTCAGAAACGCTAGACGATTTATGTCGTAAAAATGATATGGCGGTTCTTTTTCGACG 480
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 481 GGTACGAAAGGCGCGAATGGCGGCAATCAGTCTCTGTACCGGAAGATGCTATC 540
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 541 ACAGTAGGGGCAACAGAAACCTGGCGCAAGCTTTGGATCTTATGCTGACACATCAAT 600
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrIlyAspGlyArgIleLysProAspVal 220
Db 601 CATGTAGCACAGTTTTCTCCCGCGGACCTCAAGAGGATGGACGCTATCAAAACCGGACGTA 660
Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 661 ATGGCACCAGGAACATTTATTTATTCGCAAGATCTTCTTTGGCTCCGGACTCCTCATTC 720
Qy 241 TrpAlaAsnHisAspSerIlyThrAlaTyMetGlyGlyThrSerMetAlaThrProIle 260
Db 721 TGGGCAAAACCAATGACAGAAATATCTTATATGGGTGGAAATCATCGGACACCGGAT 780
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 781 GTAGTGGTAACGTTGCAGATTACGTGACATTTTCATCAAAACACAGAGGAATCACTCCT 840
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 841 AAACCATCTCTGCTGAAGACACTCTTATTCGCGGACCACTGATATCGGCTTGGCTAT 900
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 901 CCGAGTGGAAACCAAGGATGGGAGAGTAACTTGGACAGTCACTTAATGTAGCTTTC 960
Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 961 GTAAATGAACAAAGCTCTTTATCTACTAACCAAAAGGCTAGCTATTCATTTTACTGCACA 1020
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
Db 1021 TCAGGCAACACCTTTGAAGATTTTCATTTGGTTGGTCTGTATGCACCGGCAAGTACTCCGCA 1080
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db 1081 TCGGTTACATTGGTGAATGATCTGGATCTGGTGTATTACAGCTCCAAATGGAAACAAAGTAT 1140
Qy 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400
Db 1141 GTTGAACAGCTTTACTCTCTCCATGATAATACTGGATGGACGTAAACAATGTAGAG 1200
Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1201 AACGTGTTTATCAATGCTCCGCAAGCGAAGCGTATACAGTTGAGGTACAGCTTACAAAT 1260
Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1261 GTTCCACAAAGGCGCGAGCGCTTTCTTTGGCTATTGTGAAC 1302

RESULT 9

AR069954
LOCUS 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3003)
AUTHORS Sloma, A. and Christianson, L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES Location/Qualifiers
1..3003
source /organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:

Pred. No.: 8,41e-128 Length: 3003
Score: 2118.50 Matches: 405
Percent Similarity: 97.47% Conservative: 18
Best Local Similarity: 93.32% Mismatches: 10
Query Match: 94.37% Indels: 1
DB: 6 Gaps: 1

US-09-985-689A-2 (1-434) x AR069954 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCGCTGCAATGTGAAGCAGACGTCGCACAAATAAATTCCTTGCCTATAT 1529
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGAACAGGACAGATTTAGCAGTTGCTGATCTGGCTTTGTACAGAAAGAAATGACAGT 1589
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCATGCTGATGAGCATTCGCGGTAGATATCCGCATATATGCTGCTGGCAGAACGAT 1649
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAACCCCATGTTGCTGGATCTCTGTAGAAAT--- 1706
Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTACAAATAAGGATGGCACCGCAAGCCATCTAGTCTTCATCTATTATGGATAGT 1766
Qy 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 1767 GGTGGAGGGCTGGGAGGACTACCTGCTAATCTACAAACATTATTTCAGTCAAGCATATAGT 1826
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValaTyThrThr 140
Db 1827 GCTGGAGCGAGAATTATACGAATTCATGGGGGCTCCAGTAAACGCTGCTATACGACA 1886
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTCTCGAATCTTGATGATTATGTGAGAAAAAATGATATGACGATCTTTTGGCGGC 1946
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAATGTAGGAGCCCGGTAGCGGTACAATCAGTGCACCGAGGACACGAAATAATGGCAT 2006
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGCAACCCGAAACCTACGTCCAAAGCTTCGGATCTTATGCGGATATATTAAC 2066
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTGCTCAATTTCTTCCAGAGGCTCTTACAGATGGAGTATTAAGCCGCGGACGTC 2126
Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 2127 ATGGCACCAGGTACGTATATTTCTCTGCTAGATCATCATTAGTCTCCAGATTCCTCATTC 2186
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

2127	Db	TGGGCAACCATGATAGTAATAATATGCTACATGGGTGGTACTTCTATGGCTACTCAATT	2246
261	Qy	ValAlaGlyAsnValAlaGlnLeuAArgLysHisPheValIysAsnAArgGlyIleThrPro	280
2247	Db	GTAGCAGGTAATGTTGCACAAATTAAGGGAGCATTTTGTGAATAATAGAGGGTAACCTCT	2306
281	Qy	LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr	300
2307	Db	AAGCCTTCCTTTTAAAGCTGCTTAATTTGCAAGGTGCTGCGGATGTGGACTTTGGCTTT	2366
301	Qy	ProAsnGlyAsnGlnGlyTrpGlyAArgValThrLeuAspLysSerLeuAsnValAlaTyr	320
2367	Db	CCAAATGGTAACCAAGATGGGGAAGAGTAACGTTAGATAAATCCCTAAATGTCGCATTT	2426
321	Qy	ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr	340
2427	Db	GTGAATGAACAGAGCCCTTTATCAACAAGTCAAAAAGCAACATATTCGTTTACGGCTCAA	2486
341	Qy	AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla	360
2487	Db	GCTGGTAACCCCTTAAAAATATCACTTGTTTGTTCAGATGCACAGGTAGCAGCAGGCA	2546
361	Qy	SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr	380
2547	Db	TCACTAACTTTAGTGAATGATTTAGACTTTAGTAATCACTGCACAAATGAAGTAATATC	2606
381	Qy	ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu	400
2607	Db	GTCCGAAATGACTTTTACAGCACCGCTATGATAACAATTTGGGATGGCAGAAACAACGTGAA	2666
401	Qy	AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn	420
2667	Db	AATGTGTTTATCAATGTCTCTCAAGAGCGAACGTTATACGTGCAAGTGCAGGCTTACAT	2726
421	Qy	ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn	434
2727	Db	GTACACGTAAGTCGCAACCTTTTCTTAGCGAATGTACAT	2768

RESULT 10					
ED062155					
LOCUS	BD062155	3003 bp	DNA	linear	PAT 27-AUG-2002
DEFINITION	Nucleic acids encoding a polypeptide having protease activity.				

ACCESSION	BD062155
VERSION	BD062155.1
KEYWORDS	GI-22607760
SOURCE	JP 2001514529-A/39.
ORGANISM	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
1 (bases 1 to 3003)
Sloma, A. and Christianson, L.
AUTHORS
Nucleic acids encoding a polypeptide having protease activity
TITLE
Patent: JP 2001514529-A 39 11-SEP-2001;
JOURNAL

NOVO	NORDISK BIOTECH INC	COMMENT
PN	JP 2001514529-A/39	
PD	11-SEP-2001	
PF	09-JUN-1998 JP 1999503145	
PI	12-JUN-1997 US 08/873479	
PI	ALAN SILOMA,LYNNE CHRISTIANSON	
PC	C12N15/57,C12N15/75,C12N9/54,C12K14/00	
CC	Strandedness: Single;	
CC	Topology: Linear;	
EF	Key location/Qualifiers.	

FEATURES	source
Location/Qualifiers	
1..3003	
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	

ORIGIN	
Alignment Scores:	
Pred. No.:	8.41e-128
Score:	2118.50
Length:	3003
Matches:	405

Percent Similarity:	97.47%	Conservative:	18
Best Local Similarity:	93.32%	Mismatches:	10
Query Match:	94.37%	Indels:	1
DB:	6	Gaps:	1
US-09-985-689A-2 (1-434) x BD062155 (1-3003)			
Qy	1	AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr	20
Db	1470	AAYGAGTGGCCCGTGGCATTTGGAAGCAGACGTGCGCAAAATAACTTTGGCTTATAT	1529
Qy	21	GlyGlnGlyGlnIleValAlaValaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40
Db	1530	GGRCNAGGACAGATTGTAGCAGTTGCTGATPACTGGCTTGATACAGAAAGAAATGACAGT	1589
Qy	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60
Db	1590	TCGATGTCATGAAGCATTTCCCGGTAGATTACCGCACTATATGCACTGGCGCAAGCAAT	1649
Qy	61	AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly	80
Db	1650	AACGCCAATGATCCAAATGGACATGGNACCCATGTTGCTGGATCTGTGTAGAAAT--	1706
Qy	81	AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	100
Db	1707	GCTACAAATAAAGGATGGCAGCGCAAGCAATCTAGTCTTTCAATCTATTATGGATAGT	1766
Qy	101	SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer	120
Db	1767	GGTGGAGGGCTGGAGAGCACTCCTGCTTAATCTACAAACATTTTCAGTCAAGCATATAGT	1826
Qy	121	AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValaAsnGlyAlaTyrThrThr	140
Db	1827	GCTGGACGAGAAATTCATACGAATTCATGGGGGCTCCAGTAAACGGTGCTATACGACA	1886
Qy	141	AspSerArgAsnValaAspAspTyrValaArgLysAsnAspMetThrIleLeuPheAlaAla	160
Db	1887	GACTCTCGAATGTGTGATGATTATGTGAGAAAATGATATGACGATCTTTTTCGGGCC	1946
Qy	161	GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	180
Db	1947	GGAAATGAGGGACGAGTAGCGGTACAAATCAGTGCACCGAAACAGCAAAAATGCGATT	2006
Qy	181	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn	200
Db	2007	ACAGTTGGGGCAACCGAAAAACCTACGCCAAGCTTCGGATCTTATCGGATAATATTAA	2066
Qy	201	HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal	220
Db	2067	CATGTTGCTCAATTTCTTTTCACGAGGTCCTACAGATGGACGTTAATAGCGGACGTC	2126
Qy	221	MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerIleuAlaProAspSerSerPhe	240
Db	2127	ATGGCACCGAGTACGTATATCTCTCTGCTAGATCATCATTTAGCTCCAGATTTCTCATTC	2186
Qy	241	TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle	260
Db	2187	TGGGCAAACCATGATAGTAATATGCCCTACATGGGTGGTACTTCTATGGCTACTCCAA	2246
Qy	261	ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro	280
Db	2247	GTAGCAGGTAATGTTGCACAAATTAAGGAGCATTTTGTGAAAAATAGAGGGGTAACTCCT	2306
Qy	281	LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr	300
Db	2307	AAGCCTTCCCTTTTAAAGCTGCTTTAATTGCAAGTGTGGGATGTTGGACTTGGCTTT	2366
Qy	301	ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr	320
Db	2367	CCAAATGTTACCAAGGATGGGAAGAGTAACTGATTAGATAAATCCCTAAATGTCGCATT	2426
Qy	321	ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr	340


```

Db      2427 GTGAATGAAACGAGCCCTTTATCAACAAAGTCACAAAGCAACATATTCGTTACCGGCTCAA 2486
Qy      341 AlaGlyLysProLeuLysLeuValTTPSerAspAlaProAlaSerThrThrAla 360
Db      2487 GCTGGTAACCCCTTAAATAATACCTGTTGGTCAGATGCCAGGTCAGCAGCGCA 2546
Qy      361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db      2547 TCACTAACTTTAGTGAATGATTAGACTTAGTAATCACTGCACCAATGGAACATAATAC 2606
Qy      381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTTPAspGlyArgAsnValGlu 400
Db      2607 GTCGGAATAGCTTTACGACCGGTATGATACCAATGGGATGGCAGAAACAACTGGAA 2666
Qy      401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db      2667 AATGGTTTATCAATGCTCTCAAGCGGAAGCTATACAGTCAGTAGTCAGGCTTACAT 2726
Qy      421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db      2727 GTACCAGTAAGTCGCAAAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 11
AB046405 1299 bp DNA linear BCT 23-JAN-2001
LOCUS Bacillus sp. SD521 PROD gene for protease, partial cds.
DEFINITION AB046405.1 GI:12381942
VERSION AB046405.1
KEYWORDS
SOURCE Bacillus sp. SD521
ORGANISM Bacillus sp. SD521
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (sites)
AUTHORS Saeiki,K.; Okuda,M.; Hatada,Y.; Kobayashi,T.; Ito,S.; Takami,H. and Horikoshi,K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus sp.: enzymatic properties, sequences, and evolutionary relationships
JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE 20568675
PUBMED 11118284
REFERENCE 2 (bases 1 to 1299)
AUTHORS Saeiki,K.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeiki, Kao corporation, Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail:387185@kastanet.kao.co.jp, Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
     source
       1..1299
         /organism="Bacillus sp. SD521"
         /mol_type="genomic DNA"
         /strain="SD521"
         /db_xref="taxon:133780"
       1..1299
         /gene="PROD"
         <1..>1299
         /gene="PROD"
         /codon_start=1
         /transl_table=11
         /product="protease"
         /protein_id="BAB21268.1"
         /db_xref="GI:12381943"
         /translation="NDVARGIVKADVQNNYGLYGGQGVAVADTGLDTRNDSMHE
         APRGKITALYALGRNTNNDPNHGTHVAGSLGNALNKGMAPQANLVFQIMDSGG
         LGCLPSNLNTLPSQAWNAGARIHTNSWGPVNGVATNSROVDYVRNDMTVLFAAG
         NEGPNSTISAPGAKNAITVGCATENVPSPGSLADPNHIAQSPSSGATDGRKPD
         VTAPGTFILSRSSLAIPSSWPANNYSKYAMGTGSMATPIVAGNVAQLREHFKNKG
         ITPKPSLKAALAGATVGLGYPGSDGQGWGRVLDKSLNVAIYNEATALATGQKRY
         SFQQAQKPLKISLWTPDAPGSTASYTLSDLVITPAQNGQKIVGNDFSPYDNNW
         DGRNVENVFINAQSGTYTIEVQAYNVPSGPQRFSLAIVH"
     ORIGIN

```

```

Alignment Scores: 2,88e-121 Length: 1299
Pred. No.: 2010.50 Matches: 384
Score: 94.47% Conservatives: 26
Percent Similarity: 88.48% Mismatches: 23
Best Local Similarity: 89.55% Indels: 1
Query Match: 1 Gaps: 1
DB: 1
US-09-985-689A-2 (1-434) x AB046405 (1-1299)

Qy      1 AsnAspValAlaAlaArgGlyLysIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db      1 AATGATGTAGCAAGAGGAATAGTAAAGCAGACGTTGCAACAAACAAATTTACCGACTATAT 60
Qy      21 GlyGlnGlyClnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db      61 GGACAGGTCACAGTAGCTTCAGTAGCGGATACGGGTTTAGATACAGGTCGTACGATAGT 120
Qy      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db      121 TCTATGTCATGAAGCATTCGCTGGGAAAATTAACAGCTCTTTACCGGTTAGGAAGAACTAAC 180
Qy      61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db      181 AATGCAATGATCCGAATGGCAATGGTACGATGTAGCAGGTTCTGACTTGGTAAT--- 237
Qy      81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db      238 GCTTTAAATAAAGGAATGGCTCCGAAGCTAACTAGTCTTCAATCTCTATTATGGATAGC 297
Qy      101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db      298 AGCGGAGGATTTAGTGATTTACCATCGAATTTGAATACGTTATTTAGTCNAGCTTGGAA 357
Qy      121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db      358 GCTGGGGCTAGAAATTCATACTAACTCTTGGGGTGTCCAGTAATAGGAGCGGTACACTGCT 417
Qy      141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db      418 AACTCGAGCACAAGTGGATGATGTGTGAAATAATGATATGACGCTACTTTTGGACGA 477
Qy      161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db      478 GGTAATAGAGGTCCTTAATTCAGGAACAAATTAGTGCTCCAGGCACACGCAAAATGCAATT 537
Qy      181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db      538 ACGGTCGGCGCAACGGAAAACCTATCCCGGAGCTTCGGTTCATTAGCAGATAACCCAAAT 597
Qy      201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db      598 CATATCGCACAATTTTCATCAAGAGAGCTACGAGAGATGGAGCAATTAACACGACGTA 657
Qy      221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db      658 ACAGCTCCTGGAACATTTATTTATCAGCACGTCCTTCCTCTTAGCCCCAGACTCTTCGTTT 717
Qy      241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db      718 TGGCGCAATTAACAGTAGTAGTATGGCTACATGGCGGCTACTCTATGCGGACACCTATA 777
Qy      261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db      778 GTTTCGCGGGAATGTCGCGCAATTAACGAGCAATTTATAAAAAATAGAGAAATTAACCT 837
Qy      281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAspValGlyLeuGlyTyr 300
Db      838 AAACCTTCCTTAATAAAGCTGACCTTATCGCTGGGGCTACTCGATGTTGGTTAGGATAT 897
Qy      301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db      898 CCAAGTGTGTGACCAACGCTGGGGCGCTGTTACTCTAGATAAATCGTTAAATGTAGCGTAT 957

```


Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnValaThrThrThrPheThrAlaThr 340
Db 958 GTCAATGAACAACACTGCTATTAGCAACAGGCAAAAGACCAACGTAATTCGTTCOAAGCACA 1017
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
Db 1018 GCGGGTAACCTTTAAATCTCGTTAGTAGGACAGATGACCTGCGAAGTACACTGCA 1077
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db 1078 TCTATACACTAGTAAATGATTTAGATCTAGTTATTACTGCTCCGAATGGACAAAATAT 1137
Qy 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTyrAspGlyArgAsnValGlu 400
Db 1138 GTAGGAATGATTTAGTTATCTTATGATAATAAATCGGATGCTGCGAACATGTTGAG 1197
Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1198 AAGCTATTTATAACGCTCCGCAATCTGGAACGTATACAAATGAGTTCAAGCGTATAAT 1257
Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1258 GTACCACTGGCCACAGCGTTCTCACTAGCTATCGTACAT 1299

RESULT 12

AB046402
LOCUS AB046402 1299 bp DNA linear BCT 16-JUN-2001
DEFINITION Bacillus sp. D6 PROA gene for protease, partial cds.
ACCESSION AB046402
VERSION AB046402.1 GI:12381936

KEYWORDS

Bacillus sp. D6

SOURCE

Bacillus sp. D6

REFERENCE

1 (sites)

AUTHORS

Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and

Horikoshi, K.

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic Bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

20568675

11118284

2 (bases 1 to 1299)

Saeki, K.

Direct Submission

Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kasanet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Location/Qualifiers

1. 1299

/organism="Bacillus sp. D6"

/mol_type="genomic DNA"

/strain="D6"

/db_xref="taxon:127889"

1. 1299

/gene="PROA"

<1. >1299

/gene="PROA"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAB21265.1"

/db_xref="GI:12381937"

/translation="NDVARGIVKADVANNVGLVGQGVVAVADTGLDTRNDSSMHE

AFRKITALYALGRTNNANDNGHTVAGSVLGNALNKGMAQNALVFOISINDSSGG

LGLPSNINTLPSQAWNAGARIHTNSWGPVNGAYITANSQVDEIVSRNNMTVILFAPG

NEGNPSNTISAPGAKNAITVANTENRPPSGVSIADNPNIHAFSSRGATRDGR1KPD

VTAFGFTLSARSLDPSFWANYSKIAYMGTSNATPIVAGNVAQLREHFIKNRG

ITPKPSLTKAALIAAGATDVGPSGQGWRTLDKSLNVAVNEATALTGTQKATY

SFQTAGKPLKLSLWTDAPGSTTASVTLVNDLDTVITPNQKVGNDPSYFYDNNW

ORIGIN

Alignment Scores:

Pred. No.: 5,22e-121 Length: 1299

Score: 2005.50 Matches: 383

Percent Similarity: 94.24% Conservative: 26

Best Local Similarity: 88.25% Mismatches: 24

Query Match: 89.38% Indels: 1

DB: 1 Gaps: 1

US-09-985-689A-2 (1-434) x AB046402 (1-1299)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGTAGCAAGAGCAATAGTAAAGACAGACGTTTCACAAAACAATACCGACTATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrClyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGCAAGGTCAAGTAGTGTGCGATAGCGGATACGGGTTTAGATACAGTCTGTAAGTAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCTATGCATGAAGCATTCGGTGGGAAAATTACAGCTCTTTACGGGTTAGGAAGAACTAAC 180
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AATGCAATGATCGGAATGGCATGGTACGATAGCTGTCTCTGCTTGGTAAT--- 237
Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 238 GCTTTAATAAAGGAATGGCTCCGCAAGCTAAGTTAGTCTTCCAATCTATTATGGATAGC 297
Qy 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 298 AGCGGAGGATTAGGAGGATTACCAATCGAATTAATACGTTATTATTAGTCAAGCTTGGAA 357
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 358 GTGCGCGTAGAATTCATACTAATCTTGGGGGCCCCAGTAATATGGAGCGTACACTGCT 417
Qy 141 AspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 418 AACTCGAGACAGTGGATGAGTATGTTGGAACAATGATATGACGCGTCTTTTTCAGCT 477
Qy 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 478 GGAAATGAGGACCTTAATCTGGAACTATTAGCGCTCCAGGACAGCGAAAATGCCATT 537
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 538 ACGGTGCGGCAACCGAAAACCTACCGACCAAGTTTGTGTTCAATTCAGAGATAACCCCTAAT 597
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 598 CATATCGCACAAATTTTCATCGAGAGGAGCTACGAGAGATGGACGNAATTAACCCAGACCTA 657
Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 658 ACAGCTCCTCGGAACATTTAT 717
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 718 TGGGCGCAATTATAACAGTAAGTATCGGTATATGCGGCGGTACTCTTATGCGGACACCTATA 777
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 778 GTTGGGGGAATGTCGCAATTAACGCGACATTTTATAAAAAATAGAGGAATTAACACT 837
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 838 AAACCTTCCTTAATAAAAAAGCTGCACTTATCGTGGGGCTACTGATGTTGGTGTAGGATAT 897

```

QY 301 ProAsnGlyVasGInGlyTrrpGlyVargValThrLeuAspLysSerLeuAsnValAlaThr 320
Db 898 CCAAGTGTGACCAAGCGTGGGGGGTGTACTTTAGATAAATCTTAAATGTAGCGTAT 957
QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 958 GTCAATGAAGCAACTGCATTAACAACAGCAACAAAAGCAAGCAAGTATTCGTTCCAAACGCAA 1017
QY 341 AlaGlyLysProLeuValSerLeuValTrrpSerAspAlaProAlaSerThrThrAla 360
Db 1018 GCGGGTAACCAATTAAATCTCGTTAGTAGGACAGATGCACCTGGGAAGTACACAGCA 1077
QY 361 SerValThrLeuValAsnAspLeuValThrAlaProAsnGlyThrArgTyr 380
Db 1078 TCTTATACACTAGTTAATGATTAGATCTAGTTATTACTGCTCCGAATGGACAAAATAT 1137
QY 381 ValGlyVasAspPheSerAlaProPheAspAsnThrAspGlyVargAsnValGlu 400
Db 1138 GTAGTAATGATTTTAGTTATTCCTTATGATATAATGGGATGGCGCAAAATGTTGAG 1197
QY 401 AsnValPheLeuAsnSerProGlnSerGlyThrTyrThrThrLeuValGlnAlaTyrAsn 420
Db 1198 AACGTATTATAAACGCTCCGCAATCTGGAACGTATACAAATGAGTTCAAGCGTATPAC 1257
QY 421 ValProValGlyProGlnAsnPheSerLeuAlaLleValAsn 434
Db 1258 GTTCCATCTGGACCACAGCGTTCTCACTAGTATCGTACAT 1299

```

RESULT 13

```

AB046404
LOCUS AB046404 1299 bp DNA linear BCT 23-JAN-2001
DEFINITION Bacillus sp. Y PROC gene for protease, partial cds.
ACCESSION AB046404
VERSION AB046404.1 GI:12381940
KEYWORDS
SOURCE
ORGANISM Bacillus sp. Y
Bacillus sp. Y
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (sites)
AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
Horikoshi,K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE 20568675
PUBMED 11118284
REFERENCE 2 (bases 1 to 1299)
AUTHORS Saeki,K.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)

```

FEATURES

```

source
    location/Qualifiers
    1..1299
        /organism="Bacillus sp. Y"
        /mol_type="genomic DNA"
        /strain="Y"
        /db_xref="taxon:133779"
    1..1299
        /gene="PROC"
        <1..>1299
        /gene="PROC"
        /codon_start=1
        /evidence=experimental
        /transl_table=11
        /product="protease"
        /protein_id="BAB21267.1"
        /db_xref="GI:12381941"
        /translation="NDVARGIVKADVQNNQYVAGVGLGNALNKGWAPQANLVFQSIMDSGSG
        AFRGITLALYALGRNNAVDNPGHGTAVAGVGLGNALNKGWAPQANLVFQSIMDSGSG
        LGGLPSNLNTLPSQAWNAGARLHTNSWCAPVNGAYTANRSQVDEYVENDNTVLFPAG

```

gene

CDS

```

NEPNSGTTISAPCTAKNAITVGATENYRPSFGSIADNPNHIAQFSSRGATRDGRKPD
VTAPGFILUSARSSAPDSFWANTNSKTAINGGTSMATPIVAGNVQAUREHFKNRG
ITPKSLIKALIIAGATDVLGYPNGDQWGRVTLNLSNVAYNEATALATGQKATY
SQOAGKEPLKISLVMTDAPSTTAYTLVNDLDELVTAPNGKQKYGVNDFSPYDNNW
DGRNNVNFVFINAPOSQGTIIIEVQAYNVPSGPFQFSLAIVH"
ORIGIN
Alignment Scores:
Pred. No.: 1,48e-120 Length: 1299
Score: 1999.50 Matches: 382
Percent Similarity: 94.24% Conservatives: 27
Best Local Similarity: 88.02% Mismatches: 24
Query Match: 89.06% Indels: 1
DB: 1 Gaps: 1
US-09-985-689a-2 (1-434) x AB046404 (1-1299)

```

```

QY 1 AsnAspValAlaAargGlyLleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTTGCACAAAACAATACGGATTATAT 60
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyVargAsnAspSer 40
Db 61 GGACAAAGGTCAGTAGTTGTCAGTAGCGGACACAGGCTTAGATACAGGTCGTACGATAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCTATGCATCAGCATTCGCGGGAAAATCACAGCTCTTTACGCGTTAGGAGAACTAAT 180
QY 61 AsnAlaAsnAspPheAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyVasGly 80
Db 181 AATCGGAGTATCCGAATCGGCATGGCAGACACATGTAGCAGGTTCTGTACTTGGTAAT--- 237
QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 238 GCITTTAAATAAGGAATGCTCGCAAGCTAACTAGTCTTCCAATCTATTATGGATAGC 297
QY 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 298 AGCGGAGGATTAGTGGCTTACCATCGAACTAAATACGTTATTAGTCAAGCTTGAAT 357
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 358 GCTGGAGCAAGAAATTCATCTAACTCTTGGGAGCCCAAGTAAATGGACGCTACACTGCT 417
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 418 AACTCGAGACAAGTGGATGAATATGTCGAAAATAATGATATACCGGTACTTTTTCGAGCT 477
QY 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 478 GGTAAATGAAGTCTCTAATTCAGGAACAATAGTGCTCCAGGTACAGCGAAAATGCTATT 537
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 538 ACGGTCGGCGCAACGGAAAACCTATCCGCCAAGCTTCGGTTCGATAGCAGATACCCAAAT 597
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyVargIleLysProAspVal 220
Db 598 CATATTGCCCAATTTTCATCGAGAGAGCTACGAGGATGGACGATTAAGCTGACGTA 657
QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 658 ACAGCTCTCTGGAACATTTATTTATCAGCACGTTCTCTCTAGTCCAGACTCTTCGTTT 717
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 718 TGGGCGAATTTATACAGTAAATACGCTATATGCGCGGTCCTCCATGCGGACACCTATT 777
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 778 GTTGCAGGGAATGTCGCGCAATTACGTGAGCATTTTATATAAAAAATAGAGGTATTACTCT 837

```



```
Db 1721 CCTAGTGTGACCAAGCTGGGGCGTGTACTCTAGATAAATCGTAAATGTAGCGTAT 1780
Qv 321 ValAenGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPhrThrAlaThr 340
Db 1781 CTCATGAAGCAACTGCATTAGCCACAGCAACAAAAGCAACGATTCGTTCCAGACAA 1840
Qv 341 AlaGlyLysProLeuLysLysSerLeuValThrPheSerAspAlaProAlaSerThrAla 360
Db 1841 CGGGTAACCTTTAAATCTCGTTAGTAGGACAGATGCTCTCGAAGTACAACTGCA 1900
Qv 361 SerValThrLeuValAsnAspLeuValThrAlaProAsnGlyThrArgTyr 380
Db 1901 TCTTATACACATGTAATGATTAGTCTAGTATTACTGCTCGAATGACAAAATAT 1960
Qv 381 ValGlyAsnAspPheSerAlaProPhePheAsnAsnThrAspGlyArgAsnValGlu 400
Db 1961 GTAGAAATGATTTTATGTTATCTTATGATATTAACCTGGATGCTGCAACAATGTTGAG 2020
Qv 401 AsnValPheLeuAsnSerProGlnSerGlyThrTyrThrLeuGluValGlnAlaTyrAsn 420
Db 2021 AACGTTATTTAAAGCTCCGCAATCTGGACGATATATAATGAGTTCAAGCGTATAAT 2080
Qv 421 ValProValGlyProGlnAsnPheSerLeuAlaLeuValAsn 434
Db 2081 GTACCATCTGCCCAAGCGTTTCTCACTAGTATCGTACAT 2122

RESULT 15
AF268611/C
LOCUS
DEFINITION
  uncultured marine group II euryarchaeote EBAC37F11, BAC clone
ACCESSION
  AF268611
VERSION
  AF268611.1 GI:9664575
KEYWORDS
  .
SOURCE
  uncultured marine group II euryarchaeote 37F11
  Archaea; Euryarchaeota; Marine Group II; environmental samples.
REFERENCE
  1 (bases 1 to 60006)
  Nguyen,L.P., Villacorta,R., Amjadi,M., Garrigues,C.,
  Jovanovich,S.B., Feldman,R.A. and DeLong,E.F.
  Construction and analysis of bacterial artificial chromosome
  libraries from a marine microbial assemblage
  Environ. Microbiol. 2 (5), 516-529 (2000)
  2 (bases 1 to 60006)
  Beja,O., Suzuki,M.T., Koonin,E.V., Aravind,L., Hadd,A.,
  Nguyen,L.P., Villacorta,R., Amjadi,M., Garrigues,C.,
  Jovanovich,S.B., Feldman,R.A. and DeLong,E.F.
  Direct Submission
  Submitted (17-MAY-2000) R & D, Monterey Bay Aquarium Research
  Institute, P.O. Box 628, Moss Landing, CA 95039-0628, USA
  Location/Qualifiers
    1..60006
    /organism="uncultured marine group II euryarchaeote 37F11"
    /mol_type="genomic DNA"
    /db_xref="taxon:133822"
    /clone="BAC 37F11"
    /note="environmental sample"
    1484..3178
    /note="37F11#1"
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="S-layer protein"
    /protein_id="AAF97179.1"
    /db_xref="GI:9664576"
    /translation="MEPGQTELDIFVQSPSSATAGQVNIKIVSGSGGLEPEPI
    PVVSPSPFNLSQSFVSGGYPLANNQNTWNTSHVTIENLPVGNISGMQ
    SASLSPHARGIPFEIIPFTWQKPLIELSIYHPQGMVESITVIHNSIFASTP
    VLSSGAGSQSPSLHLSINSPSTVTVPDATNQDEYRLIEBEGINFETISNDIAIQF
    CDS
    1..60006
    /organism="uncultured marine group II euryarchaeote 37F11"
    /product="5S ribosomal RNA"
    5816..8611
    /product="23S ribosomal RNA"
    complement(888..10192)
    /note="37F11#7; contains Zn-ribbon and TPR-like repeats"
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="unknown"
    /protein_id="AAF97183.1"
    /db_xref="GI:9664580"
    /translation="MNECAVCGILYLAGEPCACGSQLRAQQQSFSDSDAALPTEVPG
    LIDAAQWYDLEGIKAPVEQEEAASSLPFGGSESTNISRLPFGISHPGEGIPFEL
    SNQDSQEEVHVQVPEVISTNLEPVVALETEQDSPPVSAVSVEIPEVLAPEAPD
    LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTSDLENEWSIESFGGSDQIFTAEEVDIE
    VVFSLEDITVHVHEDGENYPOMEKDSVYSEKSLQPFDLHPAQALMDVGTDSASRAL
    LESGPTSIGGQWRSAAASFOQIVARFQDSGAMNNYGIALLQVASVMQHSNDPIEVS
    NSATQPEAAILSRVNRINSQSPESINLSQALLISREBKALALLDMAGNESQRES
    NFNTRAAILAOGLDYDEARLLVLPQADKLASGNLLKLAL"
    complement(10248..11969)
    /note="37F11#8"
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="permease"
    /protein_id="AAF97184.1"
```

```

VNGISILPQVSASCFISDVFOELGRAPLTFETVSCVNGDLDETRLTFTTASTNGE
TIPDLGFTTIQNRSEFRNVSVDQMPPEPQGLIRVVGVDNGIEVVISIEPAQVSRQ
SNWYQIASFSATGLDIAITNTYNEVLGDVNCILITVTSSTSYIERIVDVGGQFA
PIVKQNPISIDRESLSATIDCNPSFDDSDPDNTATAIYVESQSLVTTNNLLWGT
ALITFVGAYVLVQMORREAGVLQENIRSKPKPTNNKDTQETKETEVIEDDMSLEIEE
EMLEBEPVSLVEEITVSETDLSPSGRLDITRIQELDDPVEIVDTTISIERNSKFPD"
3223..3891
/note="37F11#2"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97180.1"
/db_xref="GI:9664577"
/translation="MNQSNSTSAFYTLDGCVSENLDETFOALLIIRGNHQLWEXLV
NREASTRIUKQDEIDDRNKITWDDATLMTFCVRLNRSAPIMLGENKRRFGR
PHRTGSGYILENLSIQSEGLYSLDEHLEYLTERSEMNTGHEKFEKAGL
MILGITTFFBEVKALRSMILGSGMWVSREPLDGGVREAVRHLNALLMAERRGAGLIH
ROHE"
complement(3892..4797)
/note="37F11#3"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="SAICAR synthase"
/protein_id="AAF97181.1"
/db_xref="GI:9664578"
/translation="MGEMLYQGVKQVMTDPPDILERRFTNOISVFDQIIPSLIPRK
GSLNRTTAHFKLYEQSGVCTHVLNAPRCILVRKVVEVKEPGALPRDAEWFVP
LRFIVHYLUSGSAWRKFORGETASLGVQDAEYGVLSSEFVEVTTTIFPADRNID
REALQISNITDQFDSIVQAALKVDEIEREAKNGLIHVDGKKEFALGPGRKVLV
DTFGLDEDRMDAEAYSNGECIELSEKFEVTHYNSGHAELKLARDSGAIDPPIPA
LPESVIEETASLYSSMYERLTSGQF"
complement(4846..5508)
/note="37F11#4"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="translation initiation factor 6"
/protein_id="AAF97182.1"
/db_xref="GI:9664579"
/translation="MPLDLDHGSQDGIHILASVGNLPHPRELPTPTIEKLEOVVG
LELATISGGSNLIGALLAGNTKGNVADIATSIDALTAIGDVVMEGGVNTAGNL
LVNQGCVASPSIPADGLDILSEVLVDITSTIGGQDVSGSLGVVNSQFVLLKPDV
APEVLLIEILGVPPMVGTGVSFPGYVAGICASNNAGALASSETTGPENLRIEDALG
Li"
5597..5745
/product="5S ribosomal RNA"
5816..8611
/product="23S ribosomal RNA"
complement(888..10192)
/note="37F11#7; contains Zn-ribbon and TPR-like repeats"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97183.1"
/db_xref="GI:9664580"
/translation="MNECAVCGILYLAGEPCACGSQLRAQQQSFSDSDAALPTEVPG
LIDAAQWYDLEGIKAPVEQEEAASSLPFGGSESTNISRLPFGISHPGEGIPFEL
SNQDSQEEVHVQVPEVISTNLEPVVALETEQDSPPVSAVSVEIPEVLAPEAPD
LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTSDLENEWSIESFGGSDQIFTAEEVDIE
VVFSLEDITVHVHEDGENYPOMEKDSVYSEKSLQPFDLHPAQALMDVGTDSASRAL
LESQPTSIGGQWRSAAASFOQIVARFQDSGAMNNYGIALLQVASVMQHSNDPIEVS
NSATQPEAAILSRVNRINSQSPESINLSQALLISREBKALALLDMAGNESQRES
NFTNRAAILAOGLDYDEARLLVLPQADKLASGNLLKLAL"
complement(10248..11969)
/note="37F11#8"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="permease"
/protein_id="AAF97184.1"
```

```
/db_xref="GI:9664581"
/translation="MGNIILFSKRHPAGLIAIALISLLFVGTWQFIRSSHAAGGDDIDL
TVIRDFENQAVYVGLGILFVYALIIIEVVRHTLAQTGLLAVIALINHSYVEEAL
SLKATVTDWETIGLLGMVMVGIIISHTGVFEFAVAYKKSGGEIWTIVLWILCVS
TAVLSAFIDNVTMLLPVTIKIAQVLDLKEIPILISELVFSNGGAATWIGDPPNI
MIGSAMDPAISGNKDPNIANLANSQVTFNDPIETEMAGILMTIVPSFMLIRWLYKDE
FSGRDRDTELERQAKNDKGLKVSQAILTLVLGFFLHFVHLAVSWALSGALV
MLLANHELEPLEHEWETWTLFFAGLTVLHVSQYMGVVDIYGEYVERKAIFPPD
YRLAAALILWVSATASAFIDNIPIYATMTIPVLSLAFELNPLNPLIMLAFAGCL
GGNGLIIGASNVVTAAGEEAGYPISENFPRAGFPVMLLSTFIVSFYMLIVYVGG
EDGALTWKVLVGLISLLGIIAQYSRGRKAKSPALVDDDPDEIVSTIKSVIKTGS
KGSSEEE"
complement(12019..12282)
/notes="37Fl1#9; contains KH domain"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="RNA-binding protein"
/protein_id="AAF97186.1"
/db_xref="GI:9664583"
/translation="MSVPNHIIHQAKDTSLPITMBIGKSGLTDSVVVELEGQLSSRL
VKVNRGLYPRDELKIVNMAEKINSSVFAAGNAVFRN"
complement(12279..12614)
/notes="37Fl1#10; contains Zn-finger; conserved protein"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97185.1"
/db_xref="GI:9664582"
/translation="MGRRLNRDKKHAIWSQTLQSLLNKPNQFTPEQSRFALH
LVFSRRHRIPLSEVREIMCRNTLLRYGENATIRFNHGIQTCHSCTSVRRIPY
RHSQKVKA"
complement(12624..13274)
/notes="37Fl1#11"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="adenylate kinase"
/protein_id="AAF97187.1"
/db_xref="GI:9664584"
/translation="MDEGRQGLNIVLFGPPGAGKGTQOARMWDATGLPOVSTGDM
RAVAKGTSVGIKAKYMXGALVPDSVILDLRLDHDHDAKNGWMPDGPRTVPOA
EALSEIVEVAVLAIDVPDRIVERICGRYSGCTGTVHDFNPNQNGICGSGFTE
KRADDVEETVRLAEYHTQTSPLADFYRAAGIFHEVDGDXDIDVITDILLAVLG"
complement(13284..16691)
/notes="37Fl1#12"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97188.1"
/db_xref="GI:9664585"
/translation="MYVQLSGTSQATAVAGGSASLAREVLRVAGINKFSASLIKATL
INGAEDLGTDPIDINANGWGQIDLENSLPSSGSLVDVPODERELQAGFSLIYFSD
LDGSGDITLADAEASAAQSSRLANNLDLILAPDGSVLYNDGFSGLSTTG
GSADLNNIRIPAGATTQNGDMWTVEHSGSSQRYSIIVAAATLIPKADLTPI
GNSLPSSSPLGLVLSLAWHQTLASGFRIFQEDITSGLIYDNRSSLEGG
SLDSVSFTPTSTGVKHLRLDNTNNQPELDAINGDNIIDIDITAGGLRVI
FQNGDSFTPTQDSDAASVTMDVRNETGLSLFVIAHEGTERGEVSLTSVQBPD
PTYLILLSDSKSNQVNTGVTFISGQFENDTIVLTNLDDTSASPDGATKYAR
AGSFIVVARYONQPTVSHQRYIIEIGOVSDVVVPSGTSGVSAKPGERTGFSYR
PNTCSAQVMTSCTSSQWIMLGNSNSSLFEPLNILLQDLSMDVNVFPIRANGE
PLAGSTQVCTVTSPTSDNLNRYETVDVAVSELKAPRSLYGPDGAVGPGALAVF
ADTGLVFHNTQKGNVPLDFAVLDERNGPWAAEQDEQISLSTSLVTLGPQGS
GDVEMRLVLPETAREGNSDILTRVESPQFNTLSLVVGENGLVNLVSNGVTLIA
APVNNRDTFETIEVNSGNDLLEWSTSLAPDWSIGSYNPPTSPVLSVQASVQIA
KAPNOTASGFDQMLFVNGTNGRFAELVIEVAATSPAGISVSDETIAPLFI
PRGDSGQSITFMQENIPLSGQTVRDEGNILTRDSSTISPEITDLAIGESIE
VVGKVSIDEDSIDGEMNLVLTITDGVWIEFWADTSVSSQSSGGIFGLPAVLSPV
LVLLALGVLGYGRLLQSKMDDDGTIELVAPDAHTDADHLGTERRQALDISVNDI
ASGEVSGDEIVAAALAQSLDMPFVTKAQVPIGRPPSGLPVGLPPAGLPAGL
```

```
CDS
PPAKSVPLVLPQVTTAGPPLPPGGLPDGWTQWNNHYGBOYLQRMGLN"
complement(16790..18130)
/notes="37Fl1#13"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97189.1"
/db_xref="GI:9664586"

Alignment Scores:
Pred. No.: 1,02e-29 Length: 60006
Score: 626.50 Matches: 165
Percent Similarity: 53.76% Conservative: 78
Best Local Similarity: 36.50% Mismatches: 150
Query Match: 27.91% Indels: 60
DB: 1 Gaps: 17

US-09-985-689A-2 (1-434) x AF268611 (1-60006)
Qy 13 AlaGlnSerSerTyGlyLeuTyGlyGlnGlyGlnIleValAlaValAlaAspThrGly 32
Db 17383 GCTCCTCTCTCTTTGAATCTTGATGTTCTGGGAACGATTCAGTGTACTGGC 17324
Qy 33 LeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyIleThrAla 52
Db 17323 GTCGATATGGAT-----CATCCTGACCTCATCGACGGGTTCAGCA 17282
Qy 53 LeuTyAla---LeuGlyArgThrAsnAsnAlaAsnAspThrAsn---GlyHisGlyThr 70
Db 17281 ATTAAATCTCAGTTCGGTTTAGATAGTTCCTCAATTCATAGCAACACTGGTCTATGGACG 17222
Qy 71 HisValAlaGlySerValLeuGlyAsnGly-----AlaThrAsnLysGlyMetAla 87
Db 17221 CACGTGGCTATGACTGCTTGGCGATGTCAGCGGTGATGCAGTAGCAGCGGTGTGCT 17162
Qy 88 ProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyLeuGlyGlyLeu 107
Db 17161 CCAGAGCAATCTTGATGTCGTCTCTGATGCCGAATATAAACACAGCAAGTTCGAGTC 17102
Qy 108 ProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThr 127
Db 17101 CAAGTTCAATCTATGATCTCTCTGATGCCGAATATAAACACAGCAAGTTCGAGTC 17042
Qy 128 AsnSerTrpGlyAlaAlaValAsn---GlyAlaTyThrThrAspSerArgAsnValAsp 146
Db 17041 AATGCATGGGTCTTAATGGAATACGAGCCTACACTGCTGATCA-AGATCCACAGAT 16983
Qy 147 AspTyrValArgLysAsnAspMetThrIle---LeuPheAlaAlaGlyAsnGluGlyPro 165
Db 16982 CAATATGCTCCACACATAAATATCTCTCTCTGCTTTTCGGTGTGATGATGAGGC 16923
Qy 166 AsnGlyGlyThr---IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 184
Db 16922 ACAGTTCTTCAAAGATTTACTGCTCCAGAACACGCCAAGATGTTCTTTCATTGGCTCA 16863
Qy 185 ThrGluAsnLeuAlaGlyProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 204
Db 16862 ACGGTATAT-----GGTAGC-----GTGAGCTCC 16839
Qy 205 PheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaProGly 224
Db 16838 TATTCATCGAAGGTCCAACTCTCGATGAAGAATATAACCTGATTTAGTCGCTCCCGGT 16779
Qy 225 ThrTyrIleLeuSerAlaArgSerSerLeuAlaProAsp----- 237
Db 16778 GACGGTATCTGTTCTGACAGCAGCAGAGAACAAAGACGTGTGTTGTTCACTCGCGGT 16719
Qy 238 SerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyMetGlyGlyThrSerMetAla 257
Db 16718 ATTGGTACTCACTCAATGGTTCGTTCAATGTATATGCAACTCAGCGGAACCTTCTCAAGCA 16659
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:37:54 ; Search time 336.694 Seconds
(without alignments)
5476.111 Million cell updates/sec

Title: US-09-985-689A-2

Perfect score: 2245

Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVVPGQNFSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09985689/runat.31032004.161806.4145/app.query.fasta.1.3498
-DB=N_Geneseq.29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689@cgn.1.1238/runat.31032004.161806.4145 -NCFU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq.29Jan04:*

1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2237	99.6	1920	AAx37277	AAx37277 Bacillus
2	2197	97.9	1923	AAx37279	AAx37279 Bacillus
3	2191	97.6	1923	AAx37278	AAx37278 Bacillus
4	2118.5	94.4	3003	AAV82382	AAV82382 Bacillus
5	1998.5	94.0	1299	AAQ27516	AAQ27516 Alkali-pr
6	457.5	20.4	1977	AAx85667	AAx85667 Thermococ
7	457.5	20.4	1977	AAx05926	AAx05926 WO9856926
8	425.5	19.0	1236	AAx05920	AAx05920 Hyperther

9	425.5	19.0	1566	2	AAx85668	AAx85668 Pyrococu
10	425.5	19.0	1962	2	AAx85695	AAx85695 Pyrococu
11	425.5	19.0	1862	2	AAx05929	AAx05929 Hyperther
12	414.5	18.5	1977	2	AAx85669	AAx85669 Protease
13	383	17.1	2121	4	ABL54900	ABL54900 T. yonsei
14	354.5	15.8	2539	2	AAx61454	AAx61454 Streptomy
15	354.5	15.8	2809	2	AAx61455	AAx61455 DhpA-mel
16	343	15.3	135638	7	ABX34289	ABX34289 S. atrool
17	338.5	15.1	1329	6	ABK74643	ABK74643 Bacillus
18	316.5	14.1	3413	2	AAV72330	AAV72330 F. balust
19	311.5	13.9	1859	2	AAx85677	AAx85677 Thermococ
20	308.5	13.7	1560	7	AAx237569	AAx237569 Streptomy
21	308.5	13.7	59816	7	ABZ37516	ABZ37516 Streptomy
22	308.5	13.7	59816	7	ABZ37515	ABZ37515 Streptomy
23	307	13.7	2835	2	AAx08141	AAx08141 Hyperther
24	307	13.7	4765	2	AAx08132	AAx08132 Protease
25	307	13.7	4765	2	AAx85670	AAx85670 Pyrococu
26	307	13.7	4765	2	AAx05921	AAx05921 WO9856926
27	302.5	13.5	2532	2	AAQ29134	AAQ29134 Encodes R
28	300	13.4	898	2	AAx08131	AAx08131 Hyperther
29	291.5	13.0	564	2	AAx08134	AAx08134 DNA seque
30	291.5	13.0	564	2	AAx85676	AAx85676 Thermococ
31	283	12.6	3788	9	ADD24905	ADD24905 DNA encod
32	282	12.6	546	4	ABL53453	ABL53453 T. yonsei
33	277	12.3	3743	3	AD524901	AD524901 DNA encod
34	273	12.2	1306	6	ABL55784	ABL55784 Bacillus
35	273	12.2	1330	6	ABL55787	ABL55787 CspA codi
36	269.5	12.0	4716	9	ABQ80437	ABQ80437 Group B S
37	269.5	12.0	4740	7	ABZ58957	ABZ58957 Aquaricin
38	262.5	11.7	2273	2	AAQ04339	AAQ04339 Aquaricin
39	262.5	11.7	2274	2	AAQ12838	AAQ12838 Aquaricin
40	262.5	11.7	2274	2	AAQ75859	AAQ75859 Aquaricin
41	262.5	11.7	4850	6	ABN71526	ABN71526 Streptoco
42	262.5	11.7	4710	6	ABN71162	ABN71162 Streptoco
43	262.5	11.7	4734	6	ABN70525	ABN70525 Streptoco
44	262.5	11.7	4770	6	ABN69191	ABN69191 Streptoco
45	262.5	11.7	110000	6	ABN71527_19	Continuation (20 o

ALIGNMENTS

RESULT 1

AAx37277

ID AAx37277 standard; DNA; 1920 BP.

XX AC AAx37277;

XX AC AAx37277;

XX DT 20-MAR-2003 (revised)

XX DT 21-JUL-1999 (first entry)

XX DE Bacillus alkaline protease encoding DNA.

XX KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;

XX KW washing composition; oxidising agent; ss.

XX OS Bacillus sp.

XX PN WO9918218-A1.

XX PD 15-APR-1999.

XX PF 07-OCT-1998; 98WO-JP004528.

XX PR 07-OCT-1997; 97JP-00274570.

XX PA (KAOS) KAO CORP.

XX FI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;

XX PI Shikata S, Nomura M;

XX DR WPI; 1999-2877736/27.

XX DR P-PSDB; AAY17087, AAY17089.

PT Alkali protease from *Bacillus* used in washing powders.

XX Disclosure; Page 53-58; 71pp; Japanese.

XX The invention relates to alkaline proteases produced by strains of
CC *Bacillus*. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease encoding
CC DNA. (Updated on 20-MAR-2003 to correct DR file.)

XX Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,516-169 Length: 1920
Score: 2237.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.64% Indels: 0
DB: 2 Gaps: 0

US-09-985-689A-2 (1-434) x AAX37277 (1-1920)

QY 1 AsnAspValAlaArgGlyLeuValValAlaAspValAlaGlnSerSerTrpGlyLeuTyr 20
Db 616 AATGATGTGCCAGAGGTATTGTCAAGCGGATGTGGCAGACAGACTACCGTTTGTAT 675
QY 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGAACAGCCAGATTGTCCGAGTTGCCGATCTGGATTGGATACAGGAAGAAACACACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAAGCCATCCGCGGTAAATAACAGCAGCTATATGCACTGGGTGGCGAAT 795
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCGAATGATACGAACGGTCACTGGTACCCATGTGGCAGGTTCCGATTAGAAATGCC 855
QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAATAAGGAATGGCACCTCAAGCGAATCTGTTTTCAATCCATCATGGATAGC 915
QY 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrIleuPheSerGlnAlaPheSer 120
Db 916 AGTGGTGGGCTTGAGGCTTGCCTTCCATCTCGCAACCTTATTCAGCCAAAGCAITTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaAlaValAsnGlyAlaTyrThrThr 140
Db 976 GCAGTCTCCAGATTTCATACAACTCTCTGGGGGCGAGCGGTGAATGGGGCTACACGACA 1035
QY 141 AspSerArgAsnValAlaAspTyrValArgLysAsnAspValThrIleuPheAlaAla 160
Db 1036 GAITTCAGAAATGTGGATGACTATGTAAAGAAATGATATGACGATTTCTTCGCGGT 1095
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1096 GGGNATGAAGCGCGAGCGGGTACCATTAGTGCACCTGGTACGGCTAAACACGCCATA 1155
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTrpAlaAspAsnIleAsn 200
Db 1156 ACAGTCGCGCAACCGAAACCTCGCTCCAGGCTTCGTTCCCTATGCAGATATATTAAC 1215
QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220

Db 1216 CACGTTGCACAGTTCTCTTCCCGCCGACAAAGATGGCGGAATCAAGCCTGATGTC 1275
QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1276 ATGGCCCAAGGACATACATTTTATCAGCAAGATCTCTCTTGCACCCGATTCCTCTTC 1335
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGCAATCATGACACGCAAAATATGCTATATGGGTGGAACGCTCCATGGCAACCGGAT 1395
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1396 GTTTCGGGGAATGTTCACAGCTCCGTGAGCATTTTGTGAAAATAGAGGAATCACTCT 1455
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 1456 AAGCCTTCCCTATTGAAGCAGCTTTGATTGAGGTGCTGCTGATGTTGGATTGGTTAT 1515
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAACCGAAACCAAGGATGGGCGCGAGTGACCCCTGGATAAATCGTTGAACGTTGCCAT 1575
QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 1576 GTGAACGATCCAGTGCCTCATCACTAGCCAAAAGCGACATATACCTTTACTGCAACG 1635
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1636 GCGGGCAAGCCATTGAAAATCTCCCTGATGTGCGATGCCCTGCCCTCAAGCAGCTACTGCT 1695
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db 1696 TCTGTAAACCTGGTCAATGATTTGGATTGGCTCATTCAGCACCACCAACGNAACAAGAT 1755
QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1756 GTCGGGAATGACTCTCAGCACCATTTGACAAATACTGGGATGCCCGCAATAACGTAGA 1815
QY 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGTATTTTATTAATTCGCCCAAGGTGGAACATATACCATTTGAGGTGCAGCATATAT 1875
QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1876 GTGCGGTTGGACCAACAACTTCTCGTTGGCAATTGTGAAC 1917
RESULT 2
AAX37279
ID AAX37279 standard; DNA; 1923 BP.
XX
XX AAX37279;
XX
XX 20-MAR-2003 (revised)
DT 21-JUL-1999 (first entry)
XX
XX Bacillus alkaline protease encoding DNA.
XX
XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent; ss.
XX
XX Bacillus sp.
XX
XX WO9918218-A1.
XX
XX 15-APR-1999.
XX
XX 07-OCT-1998; 98WO-JP004528.
XX
XX 07-OCT-1997; 97JP-00274570.
XX
XX (KAOS) KAO CORP.
XX
XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
PI

PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17091.
 XX
 PT Alkali protease from *Bacillus* used in washing powders.
 XX
 ES Disclosure; Page 63-68; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,01e-165 Length: 1923
 Score: 2197.00 Matches: 420
 Percent Similarity: 99.77% Conservative: 13
 Best Local Similarity: 96.77% Mismatches: 1
 Query Match: 97.86% Indels: 0
 DB: 2 Gaps: 0
 US-09-985-689A-2 (1-434) x AAX37279 (1-1923)
 QY 1 AsnAspValAlaArgGlyValValValValAlaAlaGlnSerSerThrGlyLeuTyr 20
 DB 619 AATGATGTTGGCGTGGGAATGTCAAGCGGATGTGGCTCAGACAGCTACGGGTTGTAT 678
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 679 GGACAAGACAGATCGTAGCGTTCGCCGATACAGCGCTTGATACAGCTGCGAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 739 TCGATGTCATGAAGCCCTCCCGGGGAAATTAATCTGCATTATATGCAATGCGGACGACAA 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 799 AATGCCAATGATACGAATGGTCATGGTACGCAATGTGGCTGGCTCGATTAGGAACGGC 858
 QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 859 TCCACTAATAAGGAATCGCGCTCAGCGCAATCTAGTCTTCCCAATCTATCATGGATAGC 918
 QY 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 DB 919 GGTGGGGACTTGGAGGACTACCTTCGAATCTGCAGCAATCTTATTCAGCCAGCATACAGT 978
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 DB 979 GCTGTGCCAGAAATTCATACAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 1039 GATTCAGAAATGTGGAGTATGTGCGCAAAATATGATAGCATCCITTTTCGCTGCC 1098
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 DB 1099 GGGATGAGGACCGGACCGGACCATCAGTGCACCCAGCAGCAGCTAAATATGCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

1159 ACAGTCGGAGCTACGGAAAACTCCGCCAGAGCTTTGGTCTTATCGGAGCAATATCAAC 1218
 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 1219 CATGTGGCACATTTCTTTCAGTGGACCGCAAAAGGATGGACGGATCAAAACCGGATGTC 1278
 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
 1279 ATGGCACCGGAAACGTTCACTATACAGNAGATCTTCTTTGACCGGATTCCTCCTTC 1338
 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 1339 TGGCGCAACCATCAGCAATAATATGCATACATACATGCGTGGAAACGCTCCATGCTAC 1398
 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 1399 GTTCTTGGAAACGTGGCAGCTTCTGAGCATTCTTGAAACACAGGAGCATCACCA 1458
 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
 1459 AAGCCTTCTCTATTAAAGCGGCACTGATTGCGGCTGAGCTGACATCGGCTTGGCTAC 1518
 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 1519 CCGAACGGTACCAAGGATGGGACGAGTGACATTGGATAAATCCCTGACCTTGCCTAT 1578
 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaAlaThrTyrThrPheThrAlaThr 340
 1579 GTGACGAGTCCAGTCTCTATCCACGCAAGCAAGCGAGCTACTCGTTTACTGCTACT 1638
 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 1639 GCGCGCAAGCCTTTGAAATCTCCCTGCTATGCTGATGCGCCCTGCGAGCAACATGCT 1698
 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
 1699 TCGTAACGCTTGTCAATGATCTGGACCTTGTTCATTACCGCTCCCAATGGCACACAGTAT 1758
 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400
 1759 GTAGGAAATGACTTTACTTCGCCATACAATGATAAATCGGATGGCGCAATTAACGTAGAA 1818
 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 1819 ATGTATTATTATATGCACCAAGCGGACGTATACATTGAAGTACAGGCTTATTAAC 1878
 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 1879 GTACCGTTGGACCAAGAACTCTCGTTGGCAATTGTGAAT 1920

RESULT 3
 AAX37278
 ID AAX37278 standard; DNA; 1923 BP.
 XX
 AC AAX37278;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE *Bacillus* alkaline protease encoding DNA.
 XX
 KW Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX
 OS *Bacillus* sp.
 XX
 PN WO9918219-A1.
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX

PR 07-OCT-1997; 97JP-00274570.
 XX (KAOS) KAO CORP.
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 XX Shikata S, Nomura M;
 XX WPI, 1999-287736/27.
 DR P-PSDB; AAY17088, AAY17090.
 XX Alkali protease from *Bacillus* used in washing powders.
 PT Disclosure; Page 58-63; 71pp; Japanese.
 PS The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,03e-165 Length: 1923
 Score: 2191.00 Matches: 419
 Percent Similarity: 99.54% Conservative: 13
 Best Local Similarity: 96.54% Mismatches: 2
 Query Match: 97.59% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-2 (1-434) x AAX37278 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerThrGlyLeuTyr 20
 Db 619 AATGATGTTGGCGTGGGATTTGCAAGCGGATGTGGCTCAGACGAGTCTACGGGTGTAT 678
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAAGGACAGATCGTAGCGTTGCCGATACAGGCGCTTGATACAGGTGCGAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyValArgThrAsn 60
 Db 739 TCGATGCAATGAACCTTCGCGGGAATATCTGCATTATATGCTTGGACGGAGAT 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGCTATGTGTACGCATGTGGCTGGCTCGGCTATTAGGAACGCGC 858
 QY 81 AlaThrAsnIleGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAAGATGGCGCTCAGCGGAATCTAGTCTTCATCTCATATGATAGC 918
 QY 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 Db 919 GGTGGGGGACTTGGAGGACTACCTCGAATCTGCAAACTTATTCAGCCCAAGCATACAGT 978
 QY 121 AlaGlyValArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyValThrThr 140
 Db 979 GCTGGTGCCAGATTCATACAACTCCTGGGGAGCAGCAGTGAATGGGCTTACACACA 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgIleAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATGACGATCCTTTTCGCTGCC 1098

QY 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 180
 Db 1099 GGAATGAAGGACCGAACGCGGACCACTAGTCAGTCACGACGACAGCTAATAATGCAATA 1158
 QY 181 ThrValGlyValThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGCTACGGMAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACAAATATCAAC 1218
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleIysProAspVal 220
 Db 1219 CATGTGGCACAGTTCTTCACGTGGACCGCAAGGATGGACGATCAACCGGATGTC 1278
 QY 221 MetAlaProGlyThrThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCGGGAACGTTCACTATCAGCAAGATCTTCTCTTCACCGGATTCCTCCTTC 1338
 QY 241 TrpAlaAsnHisAspSerIleValAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGCAACCATGACAGTAATAATGATGATCACTGGGTGGACGCTCCATGGCTACCGGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
 Db 1399 GTTGCTGGAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACA 1458
 QY 281 LysProSerLeuLeuIysAlaAlaIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
 Db 1459 AAGCCTCTCTATTAAAGCGGACGTGATGCGGTGCGAGCTGACATCGCCTTGCTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerIleAsnValAlaTyr 320
 Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATGGATAAATCCCTGAACGTTGCTAT 1578
 QY 321 ValAsnGlnSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
 Db 1579 GTGACGAGTCCAGTTCTTATCCACCGCCAAAAGACGACGACTCTCGTTTACTTGCTACT 1638
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1639 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTTCGAGGACCAACTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
 Db 1699 TCGGTAAACGTTGTCATGATCTGGACCTTGTCATTACCGCTCCAATGGCACAGTAT 1758
 QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATACTGGATGGCCGCAATAACGTAGAA 1818
 QY 401 AsnValPheIleAsnSerProGlnSerGlyThrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGTATTATTAAATGACCAAAAGCGGACGATATACAAATTGAGGTACAGGCTTATAAC 1878
 QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGGTTGGACACAGACTTCTGTTGGCAATTGTGAAT 1920

RESULT 4
 AAV82382
 ID AAV82382 standard; DNA; 3003 BP.
 XX
 AC AAV82382;
 XX
 DT 12-APR-1999 (first entry)
 XX
 DE *Bacillus* JPI170 protease gene.
 XX
 KW Protease; detergent; surfactant; leather processing; debittering;
 KW flavour; ss.
 XX
 OS *Bacillus* sp.
 XX
 PH Key Location/Qualifiers
 CDS 846..2771

```

FT      /*tag= a
FT      846..944
FT      /*tag= b
FT      1470..2768
FT      /*tag= c
FN      WC9856927-A2.
XX
XX      17-DEC-1998.
XX
XX      09-JUN-1998; 98WO-US012005.
XX
XX      12-JUN-1997; 97US-00873479.
XX      (NOVO ) NOVO NORDISK BIOTECH INC.
PA
PI      Sloma A, Christianson L;
XX
XX      WPI; 1999-080908/07.
XX      P-PSDB; AAW89547.
XX
XX      Novel protease from Bacillus subtilis LC20 - useful in laundry and
XX      PT      dishwashing detergents and for leather processing.
XX
XX      Claim 11; Page 52-53; 77pp; English.
XX
XX      This nucleotide sequence encodes a novel protease (see AAW82382) of
XX      CC      Bacillus sp. JP170 (NCIB 12513). The sequence in plasmid p170BAN is
XX      CC      contained in Bacillus subtilis LC20 NRRL B-21680. The protease gene was
XX      CC      isolated from chromosomal DNA of JP170 following preparation of probes
XX      CC      based on protease N-terminal and internal peptides (see AAW89549-50),
XX      CC      screening of chromosomal libraries, isolation of the 3' end of the gene
XX      CC      by inverse PCR (see AAW82410-11), reconstruction of the 5' and 3' ends and
XX      CC      PCR amplification (see AAW82412-16). Claimed recombinant host cells can
XX      CC      be used in a method for producing the protease. The protease is used in
XX      CC      laundry and dishwashing detergents, for institutional and industrial
XX      CC      cleaning, and for leather processing, as well as for debittering and
XX      CC      enhancing the degree of hydrolysis of protein hydrolysates, for flavour
XX      CC      development through hydrolysis of proteins, degradation of undesired
XX      CC      peptides and in enzymatic synthesis of peptides. It has enhanced
XX      CC      stability towards oxidation under alkaline conditions, e.g. towards
XX      CC      bleaching agents of the peroxy type. The invention also provides mutant
XX      CC      cells in which the protease activity is diminished. Such cells can be
XX      CC      used for the production of heterologous recombinant proteins
XX
XX      SQ      Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;
XX
XX      Alignment Scores:
XX      Pred. No.: 3,14e-159 Length: 3003
XX      Score: 2118.50 Matches: 405
XX      Percent Similarity: 97.47% Conservative: 18
XX      Best Local Similarity: 93.32% Mismatches: 10
XX      Query Match: 94.37% Indels: 1
XX      DB: 2 Gaps: 1
XX
XX      US-09-985-689A-2 (1-434) x AAW82382 (1-3003)
XX
Qy      1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db      1470 AATGACGTGGCCCGTGGCAATGTGTAAAGCAGACGCTGCACAAAATACTTGGCTTATAT 1529
Qy      21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db      1530 GGACAAGCAGACAGATTTGACGAGTTGCTGATACGGGCTTGATACAGGAAGAAATGACAGT 1599
Qy      41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db      1590 TCGATGATGAGCAGATTCGCCGGGTAGATACCCCATATATGACATCGGCAGACAGCAAT 1649
Qy      61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db      1650 AACGCCAATGATCCAAATGACATGCAACCGACCATGTTGCTGATCTGTGTAGGAAT--- 1706

```

RESULT 5

AAQ27516
 ID AAQ27516 standard; DNA; 1299 BP.
 XX
 AC AAQ27516;
 XX
 DT 05-FEB-1993 (first entry)
 XX
 DE Alkali-protease Ya enzyme gene.
 XX
 KW Alkali resistance; surface active agent resistance; detergency improver;
 KW ss.
 XX
 OS Bacillus sp. Y.
 XX
 PH Key Location/Qualifiers
 FT I..1299
 FT /*tag= a
 XX
 PN JP04197182-A.
 XX
 PD 16-JUL-1992.
 XX
 PF 28-NOV-1990; 90JP-00327110.
 XX
 PR 28-NOV-1990; 90JP-00327110.
 XX
 PA (LLOY) LION CORP.
 XX
 DR WPI; 1992-288440/35.
 DR P-PSDB; AAR26274.
 XX
 CC DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 CC resistance and improves detergency.
 CC
 CC Claim 3; Page 2; 17pp; Japanese.
 CC
 CC The sequence is that of the alkali-protease Ya enzyme gene which can be
 CC used in the recombinant production of Ya enzyme. Ya enzyme is excellent
 CC in alkali resistance and surface active agent resistance and improves
 CC detergency
 XX
 SQ Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,19e-150 Length: 1299
 Score: 1998.50 Matches: 382
 Percent Similarity: 94.24% Conservative: 27
 Best Local Similarity: 88.02% Mismatches: 24
 Query Match: 89.02% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-2 (1-434) x AAQ27516 (1-1299)

Qy 1 AsnAspValAlaArgGlyIleValIleAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATGATGACAGAGGATAGTAAAGCTGATGTCACAAAACAAATTACGGATTATAT 60
 Qy 21 GlyGlnGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACAGGTCAACTAGTTGCTAGTACGACACAGCTTAGTACAGCTGTCAGCATAGT 120
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGCATGAAGCATTCGCGGGAATACACAGCTCTTTACGGCTTAGGAAGAACTAAT 180
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATGGAGTGATCCGAATGGCGATGGACACATGATGAGGTTCTGACTTGGTAAT--- 237
 Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCTTTTAAATAAGGAATGCGCTCCGCAAGCTAACTAGTCTTCCCAATCTATTATGATGATG 297

RESULT 6

AAAT85667
 ID AAAT85667 standard; DNA; 1977 BP.
 XX
 AC AAAT85667;

Qy 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 Db AGCGAGGATTAGTGCTTTACCATCGACTTAAATACGTTATTTAGTCNAGTTGGAAT 357
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTTPGlyAlaAlaValaAsnGlyAlaTyrThr 140
 Db GCTGGAGCAAGAATTCTACTAACTCTTGGGGAGCCCCAGTAAATGGAGCGTACACTGCT 417
 Qy 141 AspSerArgAsnValaAspAspTyrValArgLyAsnAspMetThrIleLeuPheAlaAla 160
 Db AACTCGAGACAGTGGATGAGTATGTTGCAAAATGATGATGACGCTACTTTTTCAGCT 477
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db GGTAAATGAAGTCTCTAATTCAGGAACAATAGTCTCCAGGTACAGCAAAAATGCTATT 537
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db ACGTCTGGCGCAACGGAAAACCTATCGCCCAAGCTTCGGTTCGATAGCAGTAAACCAAT 597
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCCTGACGTA 657
 Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db ACAGCTCCTGGACAATTTATTTATCAGCAGCTTCCTTCCTAGCTCCAGACTCTTCCTTT 717
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db TGGCGGAATTATAACAGTAAATACGCTATATGCGCGGTACCTCCATGGCGACACTATT 777
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db GTTCAGGGAATGTGGCGCAATTACGTGAGCATTTTATAAAAATAGAGGTATTACTCCT 837
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
 Db AAGCTCTTTTAAATAAAGCTGCATTCCTCGTGTCTACTGATGTTGGTTAGGATAT 897
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db CCTAGTGTGACCAAGGCTGGGGCGTGTACTCTAGATAAATCGTTAAATGTAGCGTAT 957
 Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
 Db GTCAATGAGCAACTGCTATGACACAGGACAAAACAGCAAGTATTCGTTCCAAGCACA 1017
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db GCGGTAACCTTTAAAAATCTCGTTAGTAGGACAGATGCTCTCGGAAGTACAACTGCA 1077
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
 Db TCTTTATACACTAGTTAATGATTTAGATCTAGTTATTTACTGCTCCGAATGGCAAAAATAT 1137
 Qy 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
 Db GTAGGAATGATTTTAGTTATTCCTTATGATAATAACTGGGATGCTCGCAACAATGTTGAG 1197
 Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db AACGTATTATTAAACCTCCGCAATCTGGAAACGTTATATANTAGGTTCAAGCTATAT 1257
 Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 Db GTACCACTCTGGCCCAACAGCGTTTCTCACTAGCTATCGTACAT 1299

XX AAX05926;
 AC
 XX
 DT 06-MAY-1999 (first entry)
 XX
 DE WO9856926 Seq ID 11.
 XX
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 XX additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 XX
 OS Thermococcus celer.
 XX
 PN WO9856926-A1.
 XX
 PD 17-DEC-1998.
 XX
 XX 04-JUN-1998; 98WO-JP002465.
 PF
 XX 10-JUN-1997; 97JP-00151969.
 PR
 XX (TAKI) TAKARA SHUZO CO LTD.
 PA
 XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 PI
 XX WPI; 1999-080907/07.
 DR
 XX P-PSDB; AAW94840.
 XX
 PT Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.
 XX
 PS Disclosure; Page 53-54; 82pp; Japanese.
 XX
 CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Ala-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis
 XX
 SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5,09e-27 Length: 1977
 Score: 457.50 Matches: 136
 Percent Similarity: 44.66% Conservative: 69
 Best Local Similarity: 29.63% Mismatches: 153
 Query Match: 20.38% Indels: 101
 DB: 2 Gaps: 17

US-09-985-689A-2 (1-434) x AAX05926 (1-1977)

Qy 8 VallysAlaAspValAlaGlnSerSerTy-GlyLeuTyrglyGlnGlyGlnIleValAla 27
 Db 433 ATAGGGCCGATACCGTCTGGAACTCCCTCGGTACGACGCGGTGGTGGTGGC 492
 Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 493 ATCTCGATAGGGTATAGACGCGAAC-----CACCCCGATCTGAAG 534
 Qy 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAspThrAsn 66
 Db 535 GCGAAGGTCTAGGTGCTAGCGCGGTCAACGCGCAGGTGCGACCCCTACGATGCCAG 594
 Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsnLys----- 84
 Db 595 GGACACGGAAACCGTCTGGGGTATCGTTGCCGGAAACCGCGGCTTACTCCAGTAC 654

Qy 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 101
 Db 655 ATAGCGCTCGCCCGCGCGAAGCTCGTGGCGTCAAGGTTCTCGGTGGCGGTCG 714
 Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
 Db 715 GGAAGCTCTCCACCATCATCGCGGGTGTGACTGGGTCTGCAGACAGGCAAGTAC 774
 Qy 122 GlyAlaArgIle-----HsThrAsnSer 129
 Db 775 GGGATAAGGTCATCAACCTCTCCCTCGGTCTCCAGAGCTCCGACGGAACCGACTCC 834
 Qy 130 TrpGlyAlaAlaValAsnGlyAlaTyThrThrAspSerArgAsnValAspTyrVal 149
 Db 835 CTCAGTCAGCGCTCAACACGCTGGACGCC----- 867
 Qy 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThr 169
 Db 868 -----GGTATAGTAGTCTGCGCGCGCGCAACAGCGGCGCGCAACACTACACC 918
 Qy 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
 Db 919 GTGGCTCACCCGCGCGGAGCAAGTCTATACCTCGGTGCA----- 963
 Qy 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
 Db 964 -----GTTGACAGCAACGCAACACTCGCAGCTTCTCCAGCAGGGA 1005
 Qy 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 229
 Db 1006 CCGACCGCGGAGGAGCTCAAGCGGAGTCTGCGCCCGCGCGTGTGACATCATAGCC 1055
 Qy 230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAla 249
 Db 1066 CCGCGCGCCAGC-----GGAACGAGCATGGCCACCCCGCATAAACGACTACACC 1116
 Qy 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
 Db 1117 AAGCGCTCTGACAGCAGATGCGCCCGCACGCTTTCGGCGGTGGCGGCTCATCTC 1176
 Qy 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
 Db 1177 CAGGCCAC-----CCGAGCTGGACCCCGGCAAGGTGAAG 1212
 Qy 287 AlalaLeuIleAlaGlyAla-----AlaAspValGlyLeu 298
 Db 1213 ACCGCCCTCATCGAGACCGCGCATAGTCGCCCAAGGAGATAGCGGACATCGCTAC 1272
 Qy 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
 Db 1273 GGTGGC-----GGTAGGTGAACGTCTACAAGGCGCATCAAGTAC 1311
 Qy 317 ---AsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyr 335
 Db 1312 CACGACTACGCCAAGCTCACCTTCACCGGTCTCGCGCGCAAGGAAAGCGCCACCCAC 1371
 Qy 336 ThrPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
 Db 1372 ACCITCGAGTCAGCGCGCCACCTTCGTGACCCCGCCCTCTACTCTGGAC----- 1422
 Qy 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 Db 1423 -----ACGGGTCTGAGCGACATCGACCTCTACCTCTACGACCCC 1461
 Qy 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAsnAsnTrpAspGly 395
 Db 1462 AACGGGAACGAG-----GTTGACTCTCTACACGCC 1494
 Qy 396 ArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGlu 415
 Db 1495 TACTACGGCTTCGAGAGGTCTGGTCTACTACAACCGCGCGGAACTGACGCGTCAAG 1554

QY 416 ValGlnAlaTyrAsnValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 Db 1555 GTCGACGTACAG-----GCGCGGCGAACTACCAGGTCCGACGTCGTCAGC 1602

RESULT 8

ID AAX05920 standard; DNA; 1236 BP.
 AC AAX05920;
 DT 06-MAY-1999 (first entry)
 DE Hyperthermostable protease fragment encoding DNA.
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 OS Pyrococcus furiosus.
 PN W09856926-A1.
 PD 17-DEC-1998.
 PF 04-JUN-1998; 98WO-JF002465.
 PR 10-JUN-1997; 97JP-00151969.
 PA (TAKI) TAKARA SHUZO CO LTD.
 PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 XX WPI; 1999-080907/07.
 DR P-PSDB; AAM94836.
 XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.
 XX Claim 6; Page 37-38; 82pp; Japanese.

CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

XX SQ Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,028-24	Length:	1236
Score:	425.50	Matches:	140
Percent Similarity:	42.77%	Conservative:	61
Best Local Similarity:	29.79%	Mismatches:	147
Query Match:	18.95%	Indels:	122
DB:	2	Gaps:	19

US-09-985-689a-2 (1-434) x AAX05920 (1-1236)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
 Db 24 GTCTCAGCTCAAGTATATGGAACCTTACGTTTGGAACTTGGATATGATGTTCTCGAAT 83
 QY 24 nileValAlaValAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
 Db 84 CACATAGATATATGACACTGGATTGAC-----GCTTCTCATCC 125

QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 126 AGATCTCCAAAGGAAAGTA-----ATTGGTGGTAGATTTCGTCAATG 170
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 Db 171 TAGGAGTTATCCATACGATGACCATGGAATGCACTCATGTAGCTTCAATAGCAGCTGG 230
 QY 78 YAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGG 290
 QY 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 Db 291 APTAAAGTTCTAGGTGCGCATGGTCTCGAAGCATATCTACTATAATTAAGGAGTTGA 350
 QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAl 132
 Db 351 GTGGCGCGTTGATAACAAAGATAAGTACGGAATTAAGTTCATTAATCTTCTTGGTTC 410
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 Db 411 AAGCCAGAGCTCAGATGCTACTGACGCTCTAAAGTCAGGCTGTAAATGCGAGCTGGGATGC 470
 QY 140 zAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 Db 471 T-----GGATTAGTTGTTGTTGTTGCTCCGC 494
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
 Db 495 TGGAAACAGTGGACCTAACAGTATACAATCGTTCCTCAGCAGCTGCAAGCAAGTTAT 554
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAsnAsnIleAs 200
 Db 555 TACAGTTGGAGCC-----GTTGCAAGATATGA 581
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 582 TGTATAACAAGCTTCTCAAGCAGAGGCGCAACTGCAGACGCGCAGCTTAAGCTTGAGT 641
 QY 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 Db 642 TGTGTCTCCAGAAACTGGATAATTGCTGCCAGACCAAGT-----GGAACATGACAT 692
 QY 240 eTTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI1 260
 Db 693 GGCTCAACCAATTATGACTATTACACAGCAGCTCTCTGGACATCATATGGCACTCTCA 752
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 Db 753 CGTAGCTGTTATGACGCTCTCTGCTCCA-----GCACACCC 791
 QY 280 cLys-----ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 Db 792 GAGCTGGACTCCAGACAAGTAATAACAGCCCTCATAGNAACTGCTGATATCGTAAGCC 851
 QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
 Db 852 AGATGAATAGCCGATATAGCCTACCGTGCA-----GGTAGGCT 890
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
 Db 891 TAATGATACAAAGGCTATTAAC-----TACGATAACTATGCAAGCTAGTTCCTCCTGG 944
 QY 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
 Db 945 ATATGTGTGCAACAACAAAGCAGCAGCAAACTCACCAGTTCGTATTAGCGAGCTTCGTCT 1004
 QY 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1005 AACTGCCCATATTACTGGACAAATGCCAAT----- 1035
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385

```

Db 1036 AGCGACCTTGATCTTTACTCTACGATCCCAATGGAACACG----- 1077
QY 385 eSerAlaProPheAspAsnAsnTirAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1078 -----GTTGACTACTCTTACACCCGCTACTATGATTCGAAAGGTTGGTTATTA 1127
QY 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCCACTGATGGAACATGACACATTAAGGTTGTAACTACAGC-----GGAAG 1178

QY 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTCAAGTAGATGTGGTAAGT 1206

RESULT 9
AAT85668
ID AAT85668 standard; DNA; 1566 BP.
XX
AC AAT85668;
XX
DT 17-OCT-2003 (revised)
DT 20-APR-1998 (first entry)
XX
DE Pyrococcus furiosus protease coding sequence.
XX
KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX
OS Pyrococcus furiosus; DSM-3638.
XX
FH Key Location/Qualifiers
FT CDS 1..1566
FT FT /*tag= a
FT FT /transl_except= (pos: 1282..1284, aa: Xaa)
FT FT /note= "Xaa= Gly, Val"
XX
PN W09721823-A1.
XX
PD 19-JUN-1997.
XX
PF 07-NOV-1996; 96WO-JP003253.
XX
PR 12-DEC-1995; 95JP-00323285.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K,
PI Tsunasawa S, Kato I;
XX
DR WPI; 1997-332794/30.
DR P-PSDB; AAW24122.
XX
PT Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.
XX
PS Claim 7; Page 90-91; 159pp; Japanese.
XX
CC This sequence represents the coding sequence for the protease from
CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
CC invention. The processes of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries. (Updated on 17-
CC OCT-2003 to standardise OS field)
XX
SQ Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 1.36e-24 Length: 1566
Score: 425.50 Matches: 140
Percent Similarity: 42.77% Conservative: 61
Best Local Similarity: 29.79% Mismatches: 147
Query Match: 18.95% Indels: 122

```

```

DB: 2 Gaps: 19
US-09-985-689a-2 (1-434) x AAT85668 (1-1566)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyCl 24
Db 24 GTCTGCAGCTCAAGTTATGCAACTTACCTTTGGAACCTTGGGATATGATGGTCTCGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACATAGGAATTAATGACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAlaAsn-- 63
Db 126 AGATCTCCAAAGGAAAAGTA-----ATTGGTGGGTAGATTTTGTCTAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValIleuG1 78
Db 171 TAGGAGTTATCCATACGATGACCATGGAATGGAATGCTCAGTCTCATAGCTTCAATAGCAGCTGG 230
QY 78 YAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAAGTGGCGG 290
QY 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAGTTCTAGGTGGCGCATGGTTCTGGAAGCATATCTACTATATAATTAAAGGAGTTGA 350
QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTirpGlyAl 132
Db 351 GTGGCGCGTTGATACAAAGATAAAGTACGAATTAAGTCAATTAATCTTCTCTGGTTC 410
QY 132 a-----AlaValAsnGlyValAlaTyrThr 140
Db 411 AAGCAGAGCTCAGATGGTACTGACGCTTAACTCAGGCTGTTTAAATGCGCGGTGATGC 470
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla1 160
Db 471 T-----GGATTAGTTGTTGTTGTTGCGCG 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla1 180
Db 495 TGGAAACAGTGGACCTTAACAAGTATACAATCGGTTCTCCAGCAGCTGCAAGCAAAAGTTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAAGTATGA 591
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTTATAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGGCTTAAGCTGAGGT 641
QY 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTTGCTCCAGGAACACTGGATTAATGCTGCCAGAGCAAGT-----GGAAGTACAT 692
QY 240 eTTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI1 260
Db 693 GCGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGGCAACTCCTCA 752
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGTATTCCACGCCCTCTTGCTCCAA-----GCACACCC 791
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGACAAAGTAAAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTirpGlyArgVa 310
Db 852 AGATGAAATAGCCGATATAGCTTACGGTGCA-----GTTAGGCT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330

```


QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
 Db 1248 AGATGAATAGCCGATATAGCCACGGTGCA-----GGTAGGGT 1286
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlnSerSerAlaLeuSerThrSe 330
 Db 1287 TAATGCATACAGGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTCTACTGG 1340
 QY 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
 Db 1341 ATATGTTGCCAACAAAGCGACCCAACTCACCAGTTCGTTATTAGCGGAGCTCGTTCGT 1400
 QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1401 AACTGCCACATTACTACGGACATGCCAAT-----TACGATAACTATGCAAGCTAGTGTCTACTGG 1431
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
 Db 1432 -AGGACCTTGATCTTACTCTACGATCCCAATGGAAACAG-----TACGATAACTATGCAAGCTAGTGTCTACTGG 1473
 QY 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1474 -----GTTGACTACTCTTACACCGCCTACTATGGATTCCGAAAGGTTGGTTATTA 1523
 QY 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1524 CAACCCAACTGATGGAACATGGCAATTAAGTTGTAAGCTACAGC-----GGAAG 1574
 QY 425 cGlnAsnPheSerLeuAlaIleValAsn 434
 Db 1575 TGCAAACTATCAAGTAGATGGTAAAGT 1602

RESULT 11

ID AAX05929 standard; DNA; 1962 BP.

AC AAX05929;

DT 06-MAY-1999 (first entry)

DE Hyperthermostable protease encoding DNA.

DE Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.

OS Pyrococcus furiosus.

PN WO9856926-A1.

PD 17-DEC-1998.

PF 04-JUN-1998; 98WO-JP002465.

PR 10-JUN-1997; 97JP-00151969.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;

XX WPI; 1999-080907/07.

XX P-PSDB; AAW94841.

XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.

PS Disclosure; Page 59-60; 82pp; Japanese.

XX The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at

CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

XX SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.8e-24 Length: 1962
 Score: 425.50 Matches: 140
 Percent Similarity: 42.77% Conservative: 61
 Best Local Similarity: 29.79% Mismatches: 147
 Query Match: 122
 DB: 2 Gaps: 19

US-09-985-689A-2 (1-434) x AAX05929 (1-1962)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyVl 24
 Db 420 GTCTGCAGCTCAAGTTATGGCACTTACGTTGGAACTGGGATATGATGTTCTGGAAT 479
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
 Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 522 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTGTCATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
 Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
 QY 78 yAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGACGACGCAAGTAATGCAAGTACAGGGATGGCTCCAGAGCTAAGCTGGCGGG 686
 QY 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyLeuProSerAsnLeuGl 112
 Db 687 AATTAAGGTTCTAGGTGCGGATGGTTCTGGAAAGCATATCTACTATATAATTAAGGAGATGA 746
 QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
 Db 747 GTGGGCCGTTTGATACAAAGATAAGTACGGAATTAAGTCAATTAATCTTCTCTTGGTTC 806
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTTAAGTCAGGCTGTTAATGACGCTGGGATGC 866
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 Db 867 T-----GGATTAGTTGTTGGTGGCGC 890
 QY 160 aGlyAsnGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 891 TGGAAACAGTGGACCTTACCAATACATACGTTCTCCAGCAGCTGCAAGCAAGATTAT 950
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 TACAGTTGGAGCC-----GTTGCAAGTATGA 977
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 978 TGTATATAACAGCTTCTCAAGCAGAGGGCCAACTCGACGACGAGCTTAAGCTTACCTGAGGT 1037
 QY 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 Db 1038 TGTGTCTCCAGGAAACTGGATAATTGCTGCCAGAGCAAGT-----GGAACCTAGCAT 1088

QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
Db 1089 GGGTCAACCAATTAATGACTATACACAGCAGCTCTGGACATCAATGGCACTCTCA 1148
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeThrPr 280
Db 1149 CGTAGCTGTATTGCGCCCTCTTCTCCAA-----GCACACCC 1187
QY 280 olys-----ProSerLeuLeuLysAlaAlaLeuileAlaGlyAla----- 293
Db 1108 GAGCTGGACTCCAGACAAGTAAACAGCCCTCATAGAACTGCTATATCGTAAGCC 1247
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAenGlnGlyTyrpGlyArgVa 310
Db 1248 AGATGAATAAGCCGATATAGCTTACGGTGCA-----GGTAGGGT 1286
QY 310 lThrLeuAspLysSerLysValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
Db 1287 TAATGCATACAAAGGTATTAAC-----TACGATAACTATGCAAGCTAGTGTTCACCTGG 1340
QY 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAAACAAAGGCGCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTCGT 1400
QY 345 uLysrleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCCATATATCTGCGGCAATGCCAAT----- 1431
QY 365 lAsnAspLeuAspLeuValileThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1432 -AGCAGCCTTGATCTTACCTCTACGATCCCAATGGAACACAG----- 1473
QY 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1474 -----GTGACTACTCTCTACACCCCTACTATGATTCGAAAGGTGGTTATTA 1523
QY 405 nSerProGlnSerGlyThrTyrTyrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACATGGACATTAAGTTTGAAGCTACAGC-----GGAAG 1574
QY 425 oGlnAsnPheSerLeuAlaileValAsn 434
Db 1575 TGCAAACTCAAGTAGATGTGTGAAGT 1602

RESULT 12

AAT85669

ID AAT85669 standard; DNA; 1977 BP.

XX AAT85669;

AC AAT85669;

DT 20-APR-1998 (first entry)

DE Protease coding sequence.

KW Protease; research reagent; thermal stability; ss.

XX Synthetic.

XX WO9721823-A1.

XX 19-JUN-1997.

XX 07-NOV-1996; 96WO-JP003253.

XX 12-DEC-1995; 95JP-00323285.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;

XX Tsunashawa S, Kato I;

XX WPI; 1997-332794/30.

DR P-PSDB; AAW24123.

XX

PT Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.

XX Claim 11; Page 95-97; 159pp; Japanese.

XX This sequence represents the coding sequence for a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries

XX SQ Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,37e-23 Length: 1977
Score: 414.50 Matches: 134
Percent Similarity: 43.07% Conservative: 68
Best Local Similarity: 28.57% Mismatches: 147
Query Match: 18.46% Indels: 120
DB: 2 Gaps: 19

US-09-985-689A-2 (1-434) x AAT85669 (1-1977)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24

Db 420 GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATGATGGTTCTGGAA 479

QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44

Db 480 CACAATAGTAATATTGACACTGGAATTGAC-----GCTTCTCATCC 521

QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaSer-- 63

Db 522 AGATCTCCAAGGAAAGTA-----ATTGGGTGGTATGATTGTCAATGG 566

QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78

Db 567 TAGAGTTATCCATACGATGACCATGGAATGGAATCACTCATGTAGCTTCAATAGCAGCTGG 626

QY 78 YAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94

Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAGGGAATGGCTCCAGGAGCTAAGCTGCGGG 686

QY 94 eGlnSerIleMet-----AspSerSerGlyLeuGlyGlyLeuProSerAsnLeuG 112

Db 687 AATTAAGTTCTAGGTCCGATCGTTCTGGAACATATCTACTATATTAAGGAGTTGA 746

QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgile----- 125

Db 747 GTGGCGGTTGATAACAAAGATAAGTAGCGAATTAAGTCAATTAATCTTCTTCTGTTTC 806

QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyValaTyrThr 140

Db 807 AGCCAGAGCTCCGACGGAAACCGACTCCCTCAGTCAGCGCTCAACACCGCTGGAGCGC 866

QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla 160

Db 867 C-----GGTATAGTAGTCTGCTGCGCGC 890

QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaIl 180

Db 891 CGGCACAGCGGGCCGGAACACCTACACCGCTCGCTCCCGCGCGCGAGGAGGTGAT 950

QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200

Db 951 AACGTCGGTGCA-----GTTGACAGCAACGA 977

QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220

Db 978 CAACATGCCAGCTTCTCCAGCAGGAGGACCGACCGCGAGGCTCAAGCGGAGGT 1037

QY 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240

```

1038 1038 CGTCCGCCCGCGGTGACATACATACCCCGCGCCGACG-----GGAACGAGCAT 1088
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
240 240 etrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1089 1089 GGGCACCCCGGATAACGACTACTACACCAAGCCCTCTGGAACACGATGGCCACCCGCA 1148
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
260 260 eValAlaGly-----AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeTh 279
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1149 1149 CGTTTCGGCGGTGGCGCGCTCATCTCCAGCCCGAC----- 1185
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
279 279 xProLys-----ProSerLeuLeuLysAlaAlaLeuLeAlaGlyAla----- 293
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1186 1186 -CGAGCTGGACCCCGCAAGAGTGAAGACCGCCCTCATCGAGACCGCGCATAGTCGC 1244
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
294 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyAr 309
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1245 1245 CCCCAAGAGATAGCGGACATCGCTACGTCG-----GGTAG 1283
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
309 309 gValThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSerAl 326
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1284 1284 GGTGAACGCTACAGGCCATCAAGTACGACGACTAGCGCAAGCTCACCTTCACCGGCTC 1343
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
326 326 aLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThrAlaGlyLysProLeuLys 346
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1344 1344 CGTCGCGGACAGGGAAGCGCCACCCACACCTTCGACGTCAGCGCGGCCACCTTCGTGAC 1403
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
346 346 sIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAs 366
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1404 1404 CGCCACCCCTCTACTGGGAC----- 1433
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
366 366 nAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPheSe 386
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1434 1434 CGACATCGACCTCTACTCTCTACGACCCCAACCGGAACGAG----- 1473
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
386 386 rAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnSe 406
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1474 1474 -----GTTGACTACTCTACACCGCTACTACGCTCTCGGAGAGTTCGGTACTACAA 1526
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
406 406 rProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyProGl 426
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1527 1527 CCCGACCGCGGAACCTGGAGCGTCAAGGTCTGTCAGCTACAG-----GGCGCGGC 1577
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
426 426 nAsnPheSerLeuAlaIleValAsn 434
      ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
1578 1578 GAACACTACAGGTGCGAGCTGTCAGC 1602

RESULT 13
ABL54900
ID ABL54900 standard; DNA; 2121 BP.
AC ABL54900;
XX
DT 11-SEP-2003 (revised)
DT 31-MAY-2002 (first entry)
XX
DE T. yonsei subtilisin-like serine protease coding sequence.
XX
KW Subtilisin-like serine protease; ss.
XX
OS Thermoanaerobacter yonseiensis.
XX
FH Key Location/Qualifiers
FT CDS 142..1779
FT /tag=a
FT /product= "subtilisin-like serine protease"
XX
XX KR2000072141-A.
XX
XX 05-DEC-2000.
XX
XX 04-AUG-2000; 2000KR-00045411.
XX

```

```

PR 04-AUG-2000; 2000KR-00045411.
PA (KIMY/) KIM Y S.
XX
XX Chang HJ, Kim DH, Byun YR, Kim YS;
XX
XX WPI; 2001-298092/31.
XX
XX P-PSDB; ABB09483.
XX
XX New DNA sequence of thermophilic protein decomposition enzyme and protein
XX derived therefrom.
XX
XX Claim 1; Page 6; 15pp; Korean.
XX
XX This sequence represents the DNA encoding the Thermoanaerobacter yonsei
XX subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX to standardise OS field)
XX
XX SQ Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;
XX

Alignment Scores:
Pred. No.: 4,83e-21 Length: 2121
Score: 383.00 Matches: 124
Percent Similarity: 47.47% Conservative: 54
Best Local Similarity: 33.07% Mismatches: 133
Query Match: 17.06% Indels: 64
DB: 4 Gaps: 15

US-09-985-689A-2 (1-434) x ABL54900 (1-2121)
QY 6 GlyTLeValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIle 25
DB 445 GGAATCACAAAA-----GCACGGAGTGATTTGGAGTCACAGGAAAAATATAACA 495
QY 26 ValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAla 45
DB 496 ATACCAATTATTGACACAGGTATAGACGGAATACAGTTGACCTCA----- 543
QY 46 PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAlaAsn----- 63
DB 544 ---GGTGGAAAAATA-----ATAGGATGGAAAGACTTTATCACACACAAA 585
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
DB 586 ACTACACCATACGACGACATGGCCATGGAACTCACGTAGCAAGTATTGTCGAGGTACA 645
QY 80 GlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 96
DB 646 GGTGCTGGAACAGCTCTTTACAAAGCGGTGCTCTGATGCTTTGTTGGTAGGAATAAAA 705
QY 97 IleMetAspSer-----SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeu 114
DB 706 GTTTTAGATGCAATGGAAAGCGGAGCATGAGCACTGTAACGAGGAATTGACTGGGCT 765
QY 115 PheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaVal 134
DB 766 GTTCAAAATAAAGATGTATACGGAATCAAAAGTTATAAATTTAAGCTCGGCACTTCTACA 825
QY 135 AsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMet 154
DB 826 AGTTCTGATGGAACCTGACTCTACTCTATTAGCATGATAGAGCAGTAGATAGCGGTATT 885
QY 155 ThrIleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGly 174
DB 886 GTAGTAGTTGTAGCAGCAGGAAGAACTCTGGCCCTGCAAAATACACCATAGGTCCTCCCTGGT 945
QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
DB 946 GCTCGGAAAAAGCCATAACAGTCCACCAATGCGAGATGTA-----GGTGAA 993
QY 195 TyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGly 214
DB 994 CTTGGCTTTAACCTT-----GCAAGCTTTCCAGCCGCGGTCTCTACTGCTGACGGA 1044

```



```
QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMe 256
Db 1606 CATCGCCAGGAGTCCGGTGGAGGACCGCCGCTACATGACCACTCCGGCAGCTCAT 1665
QY 256 talaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGACCCCGCAGCTCGCGGGCGCGCCCTCTCTGAAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGACTGGACCTCCGCGGAAGTGAAGGGGGCGCTCACCGGCTC 1758
QY 293 aAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAs 313
Db 1759 CACCAAGGGCGGC---AAGTACACCCCGTTCGAGCAGGGTTCGGCCCGGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSer----- 328
Db 1816 CRAGGCGCTCCAGCAGACCGTGTACCGACCGCGTCTCGGTGAGCTTCGGGCTCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrThrPheThrAlaTh 340
Db 1876 GTGGCCGACACCGCAGCAGCGCGTCCACCAAGCAGCTGACCTACCGCAACTCGGCAC 1935
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----AlaPr 355
Db 1936 CCAGGAGCTCAGCTGAGTGTAGCTGCGACCGCCACCGACCCCAAGGGCAGCGGCC 1995
QY 355 oAlaSer-----ThrThrAlaSerValThrLeuValAsnAspLeuValIleTh 373
Db 1996 GCGCGGGCTTCTTACCGTGGCGCCACACG-----GTGAC 2031
QY 373 rAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAr 393
Db 2032 CTRCCCGCGGGCGGC-----AGCGCTCCGTGACATGACCGC 2070
QY 393 pAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrTh 413
Db 2071 CGACACCGCG-----CTCGCGGCACGCTGGACGGCGCTACTC 2109
QY 413 rIle----- 414
Db 2110 GCGCTAGTGTGCCACCGCGCGCGGCGGACAGCTCCGCGACGCGCGCGGTGCAGCG 2169
QY 415 -GluValGlnAlaTyrAsnValProVal 423
Db 2170 CGAGTGTGAGTGTGACGAGCTGACCGTGC 2197

RESULT 15
AAT61455 standard; DNA; 2809 BP.
XX AAT61455;
AC AAT61455;
DT 17-OCT-2003 (revised)
DT 06-OCT-1997 (first entry)
XX DhpA-mel chimeric gene.
XX asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
XX Streptomyces viridosporus.
OS Streptomyces antibioticus.
OS Chimeric.
XX Key Location/Qualifiers
FH 338..2809
FT CDS /*tag= a
FT mat_protein 338..2539
FT /*tag= b

/product= "DhpA_protein_product"
/note= "from S. viridosporus"
2540..2809
/*tag= c
/product= "Melanin"
/note= "from S. antibioticus"

WO9705243-A1.
13-FEB-1997.
30-JUL-1996; 96WO-JP002147.
31-JUL-1995; 95JP-00212975.
29-FEB-1996; 96JP-00067478.
XX (SAOC ) MERCANT CORP.
XX Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;
PI Isshiki K, Yoshioka T;
XX WPI; 1997-145682/13.
DR P-PSDB; AAW13668.
XX Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
PT on 4-substituted-1,4-di:hydro:pyridine derivatives to produce Chiral
PT derivatives useful for synthesis of cardiovascular drugs.
XX Claim 5; Page 37-43; 78pp; Japanese.
XX This sequence is a fusion gene encoding Streptomyces viridosporus dhpA
CC gene product (an asymmetric hydrolase) which acts on 4-substituted-1,4-
CC dihydropyridine derivatives, and melanin from S. antibioticus. The dhpA
CC enzyme allows the efficient conversion of 4-substituted-1,4-
CC dihydropyridine esters to chiral partially hydrolysed derivatives, for
CC use in the synthesis of cardiovascular drugs suitable for the treatment
CC of e.g. hypertension and ischaemic heart disease. (Updated on 17-OCT-2003
CC to standardise OS field)
XX Sequence 2809 BP; 473 A; 1041 C; 984 G; 311 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,27e-18 Length: 2809
Score: 354.50 Matches: 139
Percent Similarity: 41.49% Conservative: 56
Best Local Similarity: 29.57% Mismatches: 173
Query Match: 15.79% Indels: 103
DB: 2 Gaps: 17

US-09-985-689A-2 (1-434) x AAT61455 (1-2809)
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCCGTCCGGCAGATCGGCCCGCCCAAGCGGTGTCGCC---GGCTACGCGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGTCTGTGACACCGCGTGTGACACGAGC----- 1054
QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGCGCGGTGACCGCGTCCCAAGAACTTCACCGCGCGCGCC 1111
QY 62 AlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
Db 1112 GCGCGGACAGGTGGGCGCACGCCACCGTCCGCTCGATCGCGCGGCGGACGCGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGGCAAGTACAAAGCGCGTCCACCGCGCGCGGATCCTCAACGCGCAAGTGC 1231
QY 98 MetAspSerSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
```

Search completed: April 4, 2004, 01:10:34
Job time : 365.684 secs

```
Db 1232 CTCGACGACTCCGGT-----TTCCGGCAGACACTCCGGCATCTCCGCGCGGCATCGAGTGG 1285
QY 118 AlapheSerAlaGlyAlaArgile-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGGCGCGCAGGCGCGCGACGTCTCAACTAGCTGGCGGCATGACACACCGCGAG 1345
QY 137 aTyThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACCGACCCGCTGGAGCGCGG-GTCCGACAAGCTGCCCGCAGAGGGCGTCCTGTTCCG 1404
QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
Db 1405 CATCGCGCGCGCAACAGAGGCGCGAG-----TCGATCGGTTCCGCCCGCAGCGCGGA 1458
QY 177 sasAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCCCTCACCCTCGCGCGCC-----GTCGA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgIl 216
Db 1486 CGACAAGGACAAGCTCGCGACTTCTCTCCACCGGCCCGCTCCGCGCAGCGGCCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrTyThrIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCCGACGTCACCGCTCCCGCGCTGCGACATCACCGCCCTCGCGCGAGGCGCAACGA 1605
QY 236 cAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGAGGTCGCTGAGGACCGCGCGCTACATGACCATCTCCGACACGTGAT 1665
QY 256 rAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGCACCCCGCACGTCGCGCGCGCGCGCTCTCTGAAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGACTGCACCTCCGCGAAGTGAAGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAGGCGGC---AAGTACACCCCTTCGAGCAGGCTTCGGGCGCGATCCAGGCGCA 1815
QY 313 pLysSerLeuAsnValAlaTyThrValAsnGluSerSerAlaLeuSer----- 328
Db 1816 CAAGGCGCTCCAGCAGACGCTGACCGGACCCGCTCCGGTGAAGCTTCGGCGTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyThrPheThrAlaTh 340
Db 1876 GTGGCGGCACACCGCAGCAGCGGTCCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----AlaPr 355
Db 1936 CCAGGACGTCACGCTGAAGCTGACGTCACCGCCACCGCCCAAGGCGCAAGGCGGCCCC 1995
QY 355 oAlaSer-----ThrThrAlaSerValThrLeuValAsnAspLeuLeuValIleTh 373
Db 1996 GCGGGCTTCTTCAGCTGGGCGCCACACG-----GTGAC 2031
QY 373 rAlaProAsnGlyThrArgTyThrValGlyAsnAspPheSerAlaProPheAspAsnTr 393
Db 2032 CGTCCCGCGCGCGGC-----AGCGCTCCGTCGACATGACCGC 2070
QY 393 pAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyTh 413
Db 2071 CGACACCCCG-----CTCGCGCGCAGGTGGACGCGCGGTACTC 2109
QY 413 rIle----- 414
Db 2110 GCGGTACGTGTCGCGCACGGGCGGCGGACAGCGTCCGCGACGGCGCGCGGTGACGG 2169
QY 415 -GluValGlnAlaTyAsnValProVal 423
Db 2170 CGAGGTCTGAGTCTGACGCTGACCGTC 2197
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.5977 Seconds
(without alignments)
3787.066 Million cell updates/sec

Title: US-09-985-689A-2
Perfect score: 2245
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNPVPGQFSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO spoel/US09985689/runat 31032004 161807 4184/app query.fasta_1.3498
-DB=Issued Patents NA -QWTF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CGN_1 1.142 @runat 31032004 161807 4184 -NCFU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORE=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2237	99.6	1920	4	US-09-509-814A-3
2	2197	97.9	1923	4	US-09-509-814A-7
3	2191	97.6	1923	4	US-09-509-814A-5
4	2118.5	94.4	2003	2	US-08-873-479-41
5	457.5	20.4	1977	3	US-08-894-818B-2
6	457.5	20.4	1977	4	US-09-445-472-11
7	425.5	19.0	1236	4	US-09-445-472-2
8	425.5	19.0	1566	3	US-08-894-818B-4
9	425.5	19.0	1962	3	US-08-894-818B-34
10	425.5	19.0	1962	4	US-09-445-472-15
11	414.5	18.5	1977	3	US-08-894-818B-6
12	359.5	16.0	2539	3	US-09-000-016-3

13	359.5	16.0	2539	4	US-09-514-340-3	Sequence 3, Appli
14	354.5	15.8	2809	3	US-09-000-016-1	Sequence 1, Appli
15	354.5	15.8	2809	4	US-09-514-340-1	Sequence 1, Appli
16	311.5	13.9	1859	3	US-08-894-818B-15	Sequence 15, Appli
17	307	13.7	2835	1	US-08-750-532-2	Sequence 2, Appli
18	307	13.7	4765	1	US-08-750-532-8	Sequence 8, Appli
19	307	13.7	4765	3	US-08-894-818B-7	Sequence 7, Appli
20	307	13.7	4765	4	US-09-445-472-5	Sequence 5, Appli
21	302.5	13.5	2532	1	US-07-671-376C-4	Sequence 7, Appli
22	300	13.4	898	1	US-08-750-532-7	Sequence 7, Appli
23	291.5	13.0	564	1	US-08-750-532-11	Sequence 11, Appli
24	291.5	13.0	564	3	US-08-894-818B-14	Sequence 14, Appli
25	273	12.2	1306	4	US-09-966-921A-1	Sequence 1, Appli
26	273	12.2	1330	4	US-09-966-921A-5	Sequence 5, Appli
27	260.5	11.6	1194	4	US-09-228-352-3407	Sequence 3407, Ap
28	259	11.5	1628	4	US-09-634-238-212	Sequence 212, App
29	256	11.4	840	1	US-08-434-255-7	Sequence 7, Appli
30	256	11.4	840	1	US-08-459-967-7	Sequence 7, Appli
31	256	11.4	840	1	US-08-460-327-7	Sequence 7, Appli
32	256	11.4	840	1	US-08-459-871-7	Sequence 7, Appli
33	256	11.4	840	3	US-09-024-532-1	Sequence 1, Appli
34	256	11.4	840	4	US-09-104-623A-1	Sequence 1, Appli
35	256	11.4	840	4	US-09-019-532-1	Sequence 1, Appli
36	256	11.4	840	4	US-09-417-359A-1	Sequence 1, Appli
37	256	11.4	840	4	US-09-705-185-1	Sequence 1, Appli
38	256	11.4	840	4	US-09-338-746-1	Sequence 1, Appli
39	256	11.4	1110	1	US-08-434-255-5	Sequence 5, Appli
40	256	11.4	1110	1	US-08-459-967-5	Sequence 5, Appli
41	256	11.4	1110	1	US-08-460-327-5	Sequence 5, Appli
42	256	11.4	1110	1	US-08-459-871-5	Sequence 5, Appli
43	256	11.4	1191	1	US-08-434-255-3	Sequence 3, Appli
44	256	11.4	1191	1	US-08-459-967-3	Sequence 3, Appli
45	256	11.4	1191	1	US-08-460-327-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUYOSHI
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3

LENGTH: 1920
TYPE: DNA
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1920)
US-09-509-814A-3

Alignment Scores:
Pred. No.: 3 9e-224 Length: 1920
Score: 2237.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0

Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.64% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-2 (1-434) x US-09-509-814A-3 (1-1920)

QY 1 AsnAspValAlaArgGlyLeuValAlaGlnSerSerTyrGlyLeuTyr 20
DB 616 AATGATGTGGCCAGAGGATTTGTCACAGCGGATGGGCACAGACAGCTACGGTTGTAT 675

QY 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 676 GCACAAAGCCAGATTGTGCGAGTTCGCCGATATGCGATTGGATACAGGAAGAAACGACAGT 735

QY 41 SerMetHisGluAlaPheArgGlyLeuValAlaValAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 736 TCGATGTCATGAAGCCCTTCGCGGTAAATACAGACACTATATGACTGGTTCGACGCAAT 795

QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 796 AATGCGAATGATACGAACGGTTCATGCTACCCATGTGGCAGGTTTCGATTAGGAATGGC 855

QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 856 GCACGAATTAAGGAATGGCCTCAAGCGAATCTGGTTTTCATCCATCATGATGATAGC 915

QY 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
DB 916 AGTGGTGGCTTGGAGGCTTGCCTCCAACTGCAACGCTTATTTCAGCAAGCATTCAGT 975

QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAlaAsnGlyValAlaTyrThr 140
DB 976 GCAGGTGCCAGAAATCATACAACTCCCTGGGGGCGAGCGGTGAATGGGGCCTCACGACA 1035

QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1036 GATTCAGAAATGTGATGACTATGTAGGAAATATGATGACGATCTTTTCGCGCT 1095

QY 161 GlyAsnGlyLeuProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 1096 GGGAAATGAAGCCGACCGCGGTACCATCGATGCACTCGCTGGTACCGCTTAAACGCCATA 1155

QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 1156 ACAGTCGGCGCAACCGAAACCTGCTCCAGCTTCGTTCTCTTCGACAGATAATTAAC 1215

QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 1216 CAGCTTGCACAGTCTCTTCCGTCGGCGCGACAAAGATGGCGCAATCAAGCCTGATGTC 1275

QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
DB 1276 ATGGCCCGGACATACATTTATCAGCAAGATCTTCTTCGACCGGATTCCTCTTC 1335

QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
DB 1336 TGGCGCAATCATGACAGCAATATGCTATATGGGTGGAACTCCATGTCACACCGAT 1395

QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 1396 GTTTCGGGGGAAATGTTCCACAGCTCCCGTGGAGCATTTTGTGAAAAATAGAGGAATCACTCT 1455

QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
DB 1456 AAGCCTTCCCTATTGAAGCAGCTTTGATTGACGGTCTGCTGATGTTGGAATGGGTTAT 1515

QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 1516 CCGAAGCGAAACCAAGATGGGGCCGAGTACCGCTGATAATCGTTGAACGTTGCCTAT 1575

QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaAlaThrTyrThrPheThrAlaThr 340
DB 1576 GTGAACGAATCCAGTGCCTTATCACTAGCCAAAAAGCGACATATACCTTTTACTCAACG 1635

QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAsnThrThrAla 360
DB 1636 CGCGGCAAGCCATTGAAATCTCCCTGGTATGTTGGATGCGCTTCCGACGACTACTGCT 1695

QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
DB 1696 TCTGTAAACCTGGTCAATGATTGGATTGGTCAATACAGCAACCAACGAAACAGATAT 1755

QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTyrAspGlyArgAsnValGlu 400
DB 1756 GTCGGAATGACTCTCAGCACCATTTGACATATACTGGGATGGCGCAATTAACGTAGAA 1815

QY 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1816 AATGATTATTAAATTCGCCCAAGTGGACATATACCATTCAGGTGCAAGCATATAAT 1875

QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
DB 1876 GTCCCGGTTGGACCAACAACTCTCGTTGGCAATTGTGAAC 1917

RESULT 2
US-09-509-814A-7
; Sequence 7, Application US/09509814A
; Patent No. 6378227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUMI
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1923)
US-09-509-814A-7

Alignment Scores:
Pred. No.: 5,98e-220 Length: 1923
Score: 2197.00 Matches: 420
Percent Similarity: 99.77% Conservative: 13
Best Local Similarity: 96.77% Mismatches: 1
Query Match: 97.86% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-2 (1-434) x US-09-509-814A-7 (1-1923)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 619 AATGATGTGGCGCTGGAATTTGCAAGCGGATGTGGCTCAGACGAGCTACGGGTTGTAT 678

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGACAAAGACAGATCGTAGCGGTTGCCATACAGGGCTTGATACAGTCCATGACAGT 738

QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 739 TCATGTCATGAAGCCCTCCGCGGGAATAATCTGTCATTATATGTCATTGGGACGACGAAT 798
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGGTCATGGTACGCATGTGGCTCGCTCGTATTAGGAACCGGC 858
Qy 81 AlaThrAsnLysGlyWetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCCACTAATAAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAACTCTATCATGGATAGC 918
Qy 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 919 GGTGGGGACTTGGAGGACTCTTCGAATCTGCACAACTTATTTCAGCCAGCATACAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaIleThrThr 140
Db 979 GTGTGTGCCAGAAATTCATCAAACTCTGGGAGCAGCAGTGAATGGGGCTTACCAACA 1038
Qy 141 AspSerArgAsnValAspAspTrpValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTCTTTCGCTGCC 1098
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GGGAAATGAAGGACCGAACCAGCGGGAACCATCAGTCGCCAGCAGGACAGCTAAATAATGCAATA 1158
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTrpAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGCTACGGAAACCTCGCCCAACCTTTGGTCTTATGCCGACAATCAAC 1218
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCAGATCTCTTCACGTGGACCGACAAAGGATGGAGCATCAACCCGGATGTC 1278
Qy 221 MetAlaProGlyThrTrpIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGGCACCGGGAACGTCTATCTATCAGCAAGATCTCTTCCTGACCGGATCTCTCTTC 1338
Qy 241 TrpAlaAsnHisAspSerLysTrpAlaTrpMetGlyGlyThrSerMetAlaThrProIle 260
Db 1339 TGGCGCAACCATGACAGTAAATATGATACATCATGGTGGACGCTCCATGGCTACACCGATC 1398
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1399 GTTGCTGGAACGTGGCAGCTCTGTGAGCAATTTGTGAAACAAAGAGGATCAACCA 1458
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTrp 300
Db 1459 AAGCCTTCTCTATTAAACGCGACATGATTTGGCGTGACGCTGACATCGGCTTGGCTAC 1518
Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTrp 320
Db 1519 CCGAACGGTAACCAAGGATGGGACCGAGTGACATTTGGATAATCCCTGAAACGTTGCCCTAT 1578
Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTrpThrAlaThr 340
Db 1579 GTGAACGATCCAGTCTCTATCCACAGCCAAAGACGACTCTCGTTACTGCTACT 1638
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
Db 1639 GCGGCAAGCCTTGAATAATCTCCCTGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1698
Qy 361 SerValThrLeuValAsnAspLeuAlaIleThrAlaProAsnGlyThrArgTrp 380
Db 1699 TCGTAACGCTTGTCAATGATCTGGACCTTGTCAATTCAGCTTCCCAATGGCACACAGTAT 1758
Qy 381 ValGlyAsnAspPheSerAlaProPheAsnAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATACTGGATGGCGCGCAATAACGTAGAA 1818
Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTrpIleGluValGlnAlaTrpAsn 420
Db 1819 AATGTATTTTAAATGACCAAAAGCGGAGCTATACAAATGAAAGTACAGCTTATAAC 1878

Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGTTGGACACAGAACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 3

US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:

Pred. No.: 2,546-219 Length: 1923
Score: 2191.00 Matches: 419
Percent Similarity: 99.54% Conservative: 13
Best Local Similarity: 96.54% Mismatches: 2
Query Match: 97.59% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-2 (1-434) x US-09-509-814A-5 (1-1923)

Qy 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTrpGlyLeuTrp 20
Db 619 ATGATGTTCCGCGTGGAAATGTCAAAGCGGATGTGCTCAGACGAGCTACGGGTGTAT 678
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAGGACAGATCGTAGCGGTTGCCGATACAGGGCTTGATACAGGTCCCAATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTrpAlaLeuGlyArgThrAsn 60
Db 739 TCGATGTCATGAAGCCTTCCCGGGGAAATTTACTGCTATTATGCTATTGGACGAGCAAT 798
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGGTATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 858
Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCCACTAATAAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAACTCTATCATGGATAGC 918
Qy 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 919 GGTGGGGACTTGGAGGACTTACCTTCGAATCTGCAACCTTATTCAGCCAGCATACAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTrpThr 140

979 GCTGTCGAGCAATTCATCAAACTCCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
1039 GATTCAGAAATGTGATGACATATGTCGCAAAATGATATGACGATCTTTTCCTGCC 1098
161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaLe 180
1099 GGAATGAAGGACCGCAACCGGCAACCATCACTGACACAGGACACAGCTAAATAATCAATA 1158
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
1159 ACAGTCGGAGCTACGGAACCTCCGCCAAGCTTTGGTCTTATGCGGACATATCAAC 1218
201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
1219 CATGTGCGCAGTCTCTTCACGTGACCGCACAAAGGATGCGGATCAACCGGATGTC 1278
221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
1279 ATGGCAGCGGGAACGTTCACTATCAGCAAGATCTTCTTGACCGGATTCCTCTCTTC 1338
241 TrpAlaAsnHisAspSerIleTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
1339 TGGCGCAACCATGACAGTAAATATGATATGATGCGGTGGAACGTCCATGCTGCTACCGATC 1398
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
1399 GTTGTCTGGAACGTGCGACAGCTTCGTGAGCATTTTGTGAAACACAGAGCGATCACCA 1458
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
1459 AAGCCTTCTCTATTAAGGCGCACTGATTCGCGGTGCGAGTGCATCGCCTTGGCTAC 1518
301 ProAsnGlyAsnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
1519 CCAACCGTAAACAGATGGGACGAGTGACATGGATAAATCCCTGACGCTGCTAT 1578
321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
1579 GTGAACGAGTCCAGTTCTCTATCCACGACCAAAAGGACGCTACCTGTTTACTGCTACT 1638
341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProIleSerThrAla 360
1639 GCGCGCAAGCCTTTGAAATCTCCCTGGTATGCTCGATGCCCTCGGACCAACTGCT 1698
361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
1699 TCCGTACGCTTGTCAATGATCTGGACCTTGTCAATACCGTCCAAATGGCACACAGTAT 1758
381 ValGlyAsnAspPheSerAlaProPheAspAsnTrpAspGlyArgAsnAsnValGlu 400
1759 GTAGGAATGACVTTTACTTCGCCATACATGATGATGATGATGATGATGATGATGATGAT 1818
401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
1819 AATGATTTTATTAATGACCAACCAAGCGGACGCTATACATGAGGTACAGGCTTATAAC 1878
421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
1879 GTACCGGTGGACACAGCCTTCTCGTTGGCAATTTGTGAAT 1920

RESULT 4
US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lynne, Christiantson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agris, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3003 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-479-41
Alignment Scores:
Pred. No.: 1,98e-211 Length: 3003
Score: 2118.50 Matches: 405
Percent Similarity: 97.47% Conservative: 18
Best Local Similarity: 93.32% Mismatches: 10
Query Match: 94.37% Indels: 1
DB: 2
US-09-985-689A-2 (1-434) x US-08-873-479-41 (1-3003)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 1470 AATGACGTGGCCCTGGCATTGTGAAGCAGAGCTGCCAAATATACCTTTGGCTTATAT 1529
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 1530 GGACAAGACACAGATTGTAGCAGTTGCTGCTACTGGGCTTGATACAGGAGAAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 1590 TCGATGATGATGAGCATTCGCGGTGATGATGATGATGATGATGATGATGATGATGAT 1649
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 1650 AACGCCAATGATCCAAATGGACATGGAAACCATGTTCTGCTGATCTGTGTAGGAAT --- 1706
QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 1707 GCTCAAAATAAAGGATGGCAGCCCAAGCAATCTAGCTTTTCAATCTATTTGATGATGAT 1766
QY 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
DB 1767 GGTGGAGGCTGGGAGGACTACCTGCTAACTCAAAATATTATTCAGTCAAGCATATAGT 1826
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
DB 1827 GCTGAGCGAGAAATCATCAAGATTCATGCGGGCTCCAGTAAACCGGTGCTTATACGACA 1886
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1887 GACTCTGAAATGTTGATGATTATGTGAGAAAAATGATGATGATGATGATGATGATGATGAT 1946

QY 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAlaIle 180
Db 1947 GGAATAGGACCGAGGAGCGGTACATCAGTCAGCCAGGACGCAAAATGCGGATT 2006
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGCAACCGCAAACTCAGTCCAGCTTCGATCTTATCGGATAATTAAC 2066
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATCTCTTCAGAGGTCCTACTAGAGATGAGCGGTATTAAACCGGAGTC 2126
QY 221 MetAlaProGlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 2127 ATGGCACCAGGTACGTATATCTCTCTCTAGATCATCATTAGCTCCAGATTCCTCATTC 2186
QY 241 TrpAlaAsnHisAspSerLysTyAlaTyMetGlyThrSerMetAlaThrProIle 260
Db 2187 TGGGCAAAACCATGATAGTAATATCCCTACATGCTGCTGCTGCTGCTGCTGCTGCT 2246
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAtgGlyIleThrPro 280
Db 2247 GTAGCAGGTATGTTGCACATTAAGGAGGATTTTGTAAATAATAGGCGGTAACTCCT 2306
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuTy 300
Db 2307 AAGCCTTCCTTTTAAAGTCTGTTTAAATGCGAGTCTGCGGATGTTGGACTTGGCTTT 2366
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTy 320
Db 2367 CCATATGTTACCAAGATGGGAGAGATACGTTAGATAATCCCTAAATGTCGCAATT 2426
QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyThrPheThrAlaThr 340
Db 2427 GTGAATGAACAGCGCTTTATCAACAGTCAAAAGCAACATATTCGTTACGGCTCAA 2486
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 2487 GCTGTAAACCCCTTAAATAATACCTTGTGTCAGATGACACGAGGTAGCACACGGCA 2546
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrArgTy 380
Db 2547 TCACTAACCTTAGTCAATGATTTAGACTTAGTATACCTACGCAACCAATGGAATCAAT 2606
QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnTrpAspGlyArgAsnAsnValGlu 400
Db 2607 GTCGAAATGACTTTACAGCACCGGTATGATACCAATTTGGATGCGAGAAACACGTCGAA 2666
QY 401 AsnValPheIleAsnSerProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsn 420
Db 2667 AATGTTTATCATGCTCTCTCAAGCGGAAGTATACAGTCGAAGTCAGGCTTACAT 2726
QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAGTAAGTCGCAACCTTTTCTTTTAGGATTTGATAC 2768

RESULT 5

US-08-894-818B-2
; Sequence 2, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-May-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:

Pred. No.:	5,99e-38	Length:	1977
Score:	457.50	Matches:	136
Percent Similarity:	44.66%	Conservative:	69
Best Local Similarity:	29.63%	Mismatches:	153
Query Match:	20.38%	Indels:	101
DB:	3	Gaps:	17

US-09-985-689A-2 (1-434) x US-08-894-818B-2 (1-1977)

QY	8	ValLysAlaAspValAlaGlnSerSerTyGlyLeuTyGlyGlnGlyIleValAla	27
Db <td>433<td>ATAGGGCGGATACCGTCTGGAACCTCCCTCGCTACGACGGAAGCGGTGCTGTTGCC</td><td>492</td></td>	433 <td>ATAGGGCGGATACCGTCTGGAACCTCCCTCGCTACGACGGAAGCGGTGCTGTTGCC</td> <td>492</td>	ATAGGGCGGATACCGTCTGGAACCTCCCTCGCTACGACGGAAGCGGTGCTGTTGCC	492
QY <td>28<td>ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisclualaPheArg</td><td>47</td></td>	28 <td>ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisclualaPheArg</td> <td>47</td>	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisclualaPheArg	47
Db <td>493<td>ATCGTCGATACGGGTATAGACCGGAAC-----CACCCGATCTGGAAG</td><td>534</td></td>	493 <td>ATCGTCGATACGGGTATAGACCGGAAC-----CACCCGATCTGGAAG</td> <td>534</td>	ATCGTCGATACGGGTATAGACCGGAAC-----CACCCGATCTGGAAG	534
QY <td>48<td>GlyLysIleThrAlaLeuTy---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn</td><td>66</td></td>	48 <td>GlyLysIleThrAlaLeuTy---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn</td> <td>66</td>	GlyLysIleThrAlaLeuTy---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn	66
Db <td>535<td>GGCAAGGTCATAGGCTGGTACGACCGCTCAACGCGAGGTGCGACCCCTACCATGACACG</td><td>594</td></td>	535 <td>GGCAAGGTCATAGGCTGGTACGACCGCTCAACGCGAGGTGCGACCCCTACCATGACACG</td> <td>594</td>	GGCAAGGTCATAGGCTGGTACGACCGCTCAACGCGAGGTGCGACCCCTACCATGACACG	594
QY <td>67<td>GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsnLys-----</td><td>84</td></td>	67 <td>GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsnLys-----</td> <td>84</td>	GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsnLys-----	84
Db <td>595<td>GGACACGGAACCCACGTTGCGGGTATCGTTGCGGAACCGCGGATTAACCTCCAGTAC</td><td>654</td></td>	595 <td>GGACACGGAACCCACGTTGCGGGTATCGTTGCGGAACCGCGGATTAACCTCCAGTAC</td> <td>654</td>	GGACACGGAACCCACGTTGCGGGTATCGTTGCGGAACCGCGGATTAACCTCCAGTAC	654
QY <td>85<td>---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer</td><td>101</td></td>	85 <td>---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer</td> <td>101</td>	---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer	101
Db <td>655<td>ATAGGCGTCGCGCGCGGCGGAAGTCGTCGCGGTCAAGTTCTCGGTGCCGACGTTGCG</td><td>714</td></td>	655 <td>ATAGGCGTCGCGCGCGGCGGAAGTCGTCGCGGTCAAGTTCTCGGTGCCGACGTTGCG</td> <td>714</td>	ATAGGCGTCGCGCGCGGCGGAAGTCGTCGCGGTCAAGTTCTCGGTGCCGACGTTGCG	714
QY <td>102<td>GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla</td><td>121</td></td>	102 <td>GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla</td> <td>121</td>	GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla	121
Db <td>715<td>GGAAGCGTCTCCACCATCATCGCGGTGTTGATCGGTGCTGCGAGCAACAGACAAAGTAC</td><td>774</td></td>	715 <td>GGAAGCGTCTCCACCATCATCGCGGTGTTGATCGGTGCTGCGAGCAACAGACAAAGTAC</td> <td>774</td>	GGAAGCGTCTCCACCATCATCGCGGTGTTGATCGGTGCTGCGAGCAACAGACAAAGTAC	774
QY <td>122<td>GlyValaArgile-----</td><td>129</td></td>	122 <td>GlyValaArgile-----</td> <td>129</td>	GlyValaArgile-----	129
Db <td>775<td>GGGATAGGGGTATCATCACTCTCCCTCGCTCCTCCAGAGCTCCGACGGAACGACATCC</td><td>834</td></td>	775 <td>GGGATAGGGGTATCATCACTCTCCCTCGCTCCTCCAGAGCTCCGACGGAACGACATCC</td> <td>834</td>	GGGATAGGGGTATCATCACTCTCCCTCGCTCCTCCAGAGCTCCGACGGAACGACATCC	834

QY 130 TrpGlyAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrVal 149
Db 835 CTAGTCAGCCGCTCAACAGCGCTGGAGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyThr 169
Db 868 ---GATATAGTCTCGTGGCGCGGCAACAGCGCGGCGCAACCTACACCC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
Db 919 GTGGCTACCCCGCGCGGCAAGGTCATACCGTGGTGA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 964 -----GTTGACAGCAACGACATCGCCAGCTTCTCCAGCGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 229
Db 1006 CCAGCCGCGGAGGCTCAAGCGGAAGTGTGCGCCCGCGCTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAla 249
Db 1066 CCGCGCGCCAGC-----GGAACGAGCATGGGCAACCCCGATAAACGACTACTACAC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly-----AsnValAlaGlnLeu 268
Db 1117 AAGCCCTCGAACGACGATGGCCGCCCGGCGTTCGGCGGTGGCGCTCATCTC 1176
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
Db 1177 CAGGCCAC-----CCGAGTGGACCCCGGACAGGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspValGlyLeu 298
Db 1213 ACCGCCCTCATGAGACCGCGGACATAGTCGCCCGCAAGAGATAGCGGACATCGCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlyIleThrProLys----- 316
Db 1273 GGTGCG-----GGTAGGTGAACGCTACAAAGGCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyr 335
Db 1312 GACACATACGCAAGCTACCTTCACCGGCTCGTGGCGGACAGGAGGAGCGCCACCCAC 1371
QY 336 ThrPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
Db 1372 ACCTTCAGCGTCACGGCGCCACCTCTGTCACCGCCACCTCTACTGGGAC----- 1422
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
Db 1423 ---ACGGGCTCGAGGACATCGACCTCTACTCTCTACGACCC 1461
QY 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGly 395
Db 1462 AACGGGACAG-----GTTGACTACTCTACACCGCC 1494
QY 396 ArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrThrIleGlu 415
Db 1495 TACTACGGCTTCGAGAGGTGGCTACTACACCCCGGACCGCGGACCTGGAGCGTCAAG 1554
QY 416 ValGlnAlaTyrAsnValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1555 GTGTCAGCTACAG-----GGCGGGCGGAGCTACACGAGTGGCGTCTGTCAGC 1602

RESULT 6

US-09-445-472-11
; Sequence 11, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko

; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151989/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-11

Alignment Scores: 5.99e-38 1977
Pred. No.: 457.50 136
Score: 44.66% 69
Percent Similarity: 29.63% 153
Best Local Similarity: 20.38% 101
Query Match: 4 17
DB:

US-09-985-689A-2 (1-434) x US-09-445-472-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla 27
Db 433 ATAGGGCGGATACCGTCTGGAACTCCCTCGGTACACGAGCGGTGGTGGTTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCTGCGATACGGGTATAGCGGAAC-----CACCCGATCTCAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GCGAAGTTCATAGCTCGTACGACGCGCTCAACGGCAGGTCGACCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsnLys----- 84
Db 595 GGACACGAAACCCAGTTGCGGTATCGTTCGCGAACCGCGGAGCGTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 101
Db 655 ATAGCGTCGCGCCCGCGGCGAAGCTCGTGGCGTCAAGGTTCTCGGTGCGCGGTCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
Db 715 GGAACGCTCTCCACCATCATCGCGGTGTTGACTGGGTGTCAGAACAGGACCAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGATAGGGTCTACCACTCTCCCTCGGTCTCCAGAGCTCCGAGCGGACCGACTCC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
Db 835 CTCAGTCAGCGGCTCAACACGCTCGGACGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyThr 169
Db 868 -----GGTATAGTCTGCGTCCCGCGCAACAGCGGCGGCGCAACCTACAC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
Db 919 GTGGCTACCCCGCGCGGAGGTCATACCGTGGTGA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 964 -----GTTGACAGCAACGACATCGCCAGCTTCTCCAGCGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 229

Db 1006 CGAGCCGGGAGGAGGCTCAAGCCGGAAGTCGTCGCCCGCGGTTGACATCATAGCC 1065
Qy 230 AlaArgSerSerLeuAlaProAspSerPheThrAlaAsnHisAspSerLysThrAla 249
Db 1066 CGCGCGCCAGC-----CGAACCGATGGCCACCCCGATAAACGATCTACACC 1116
Qy 250 TyrMetGlyThrSerMetAlaThrProLeuValAlaGly---AsnValAlaGlnLeu 268
Db 1117 AAGGCTCTGGAACCAAGCATGCCACCCCGACCGTTTCGGCGTTGGCGCTCATCCTC 1176
Qy 269 ArgGluHisPheValLysAsnArgGlyLeuThrProLys-----ProSerLeuLeuLys 286
Db 1177 CAGGCCAC-----CCAGCTGGACCCCGACCAAGGTGAAG 1212
Qy 287 AlaAlaLeuLeuAlaGlyAla-----AlaAspValGlyLeu 298
Db 1213 ACCGCCCTCATCGAGACCGCCGACATAGTCGCCCCCAAGGATAGCGCATCGCCTAC 1272
Qy 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
Db 1273 GGTGGC-----GCTAGGGTGAACGCTCTACAGGCCCATCAGTAC 1311
Qy 317 ---AsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyr 335
Db 1312 GAGCACTACGCAAGCTCACCTTCACCGCTCCGTCGCCGACCAAGGAGCGCCACCCAC 1371
Qy 336 ThrPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
Db 1372 ACCTTCGACGTACGCGGCCACCTTCGTCGACCGCCACCCCTCTACTGGAC----- 1422
Qy 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACTACGACCC 1461
Qy 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGly 395
Db 1462 AACGGGAACGAG-----GTTGACTCTCTACACCGCC 1494
Qy 396 ArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGlu 415
Db 1495 TACTACGGCTTCGAGAGGTCCGGTACTACACCGCGACCGCGGAACTGACGGTCAAG 1554
Qy 416 ValGlnAlaTyrAsnValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1555 GTCGTCAAGTACAAG-----GGCGCGGGAACCTACCAGGTCCAGCTCGTCAGC 1602

RESULT 7

/ Sequence 2, Application US/09445472
/ Patent No. 6358726
/ GENERAL INFORMATION:
/ APPLICANT: TAKAKURA, Hikaru
/ APPLICANT: MORISHITA, Mio
/ APPLICANT: SHIMOJO, Tomoko
/ APPLICANT: ASADA, Kiyozo
/ APPLICANT: KATO, Ikunoshin
/ TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
/ FILE REFERENCE: TAKAKURA=6
/ CURRENT APPLICATION NUMBER: US/09/445,472
/ CURRENT FILING DATE: 1999-12-06
/ PRIOR APPLICATION NUMBER: 151969/1997
/ PRIOR FILING DATE: 1997-06-10
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 2
/ LENGTH: 1236
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic
US-09-445-472-2

Alignment Scores: 6.35e-35 Length: 1236
Pred. No.: 425.50 Matches: 136
Score: 42.77% Conservative: 61
Percent Similarity: 29.79% Mismatches: 147
Best Local Similarity: 18.95% Indels: 122
Query Match: 4 Gaps: 19
DB:

US-09-985-689A-2 (1-434) x US-09-445-472-2 (1-1236)

Qy 12 ValAlaGlnSerSerTyrGlyLeuTyr-----GlyGlnGly 24
Db 24 GTCGACGCTCACTTATGGCACTTACGTTGGAACTTGGGATATGATGTTCTGGAAT 83
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44
Db 84 CACATATGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 125
Qy 44 uAlaPheAcGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAGGAAGAAAGTA-----ATTGGGTGGTAGATTTGTCTCAATGG 170
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78
Db 171 TAGGAGTTATCCATACGATGACCATGGACATGGAACCTCATGTAGCTTCAATAGCAGTGG 230
Qy 78 yAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGAGCAGCAAGTAATGGCAAGTACAGGAATGGCTCCAGGAGTAAAGCTGGCGGG 290
Qy 94 eGlnSerIleMet-----AspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeuG 112
Db 291 AATTAAGTTCTAGTGCCGATGGTTCTGGAACATATCTACTATAATTAAAGGAGTTGA 350
Qy 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCGGCTTGATACAAAGATAAGTACGAATTAAGGTCAATTAATCTCTCTCTGTTTC 410
Qy 132 a-----AlaValAsnGlyValAlaTyrThr 140
Db 411 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGACGCGTGGGATGC 470
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla 160
Db 471 T-----GGATTAGTTGTTGTGGTTCGCCG 494
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180
Db 495 TGGAAACAGTGGACCTTAACAGTATACATCGTTCTCCAGCAGCTGCAAGCAAGTTAT 554
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGCAAGTATGA 581
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATTAACAGGCTTCTCAGCAGAGGGCCAACTGCAGCGGCGAGCTTAAGCTGAGGT 641
Qy 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGTCTCCAGGAACATGGATATTTCTGCCAGGCAAGT-----GGAACCTAGCAT 692
Qy 240 eTrpAlaAsnHisAspSerLysIleThrAlaTyrMetGlyThrSerMetAlaThrPro 260
Db 693 GGGTCAACCAATTAAATGACTATTACAGCAGCTCTCTGGGACATCATATGGCACTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGTATTGACGCCCTCTTGTCCAA-----GCACACCC 791
Qy 280 oLys-----ProSerLeuLeuLysAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGAGCTCCAGACAAAGTAAAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851

```
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyValGva 310
Db 852 AGATGAATAGCGGATATAGCTACGGTCA-----GGTAGGGT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
Db 891 TAATGCATCAAGCGCTATAAC-----TACGATAACTATGCAAGCTAGTGTTCACCTGG 944
QY 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCAACAAGGAGGAGCAACTCACCAGTTCGTATTAGCGGAGCTTCGTTGCT 1004
QY 345 uLysIleSerLeuValTyrSerAspAlaProLaserThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATTATCTGGGACAATGCCAAT----- 1035
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTACTCTACGATCCCAATGGAACCCAG----- 1077
QY 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1078 -----GTTGACTACTCTTACACCGCTACTATGATTCGAAAGGTTGGTTATTA 1127
QY 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCCACTGTAGTGAACATGGACAAATTAAAGTTGTAACTACAGC-----GGAAG 1178
QY 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206
```

RESULT 8

```
US-08-894-818B-4
; Sequence 4, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masenori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
```

TELECOMMUNICATION INFORMATION:

```
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; OTHER INFORMATION: /note= N at position 1283 is G or T.
US-08-894-818B-4
```

```
Alignment Scores: 9.23e-35 Length: 1566
Pred. No.: 425.50 Matches: 140
Score: 425.50 Conservative: 61
Percent Similarity: 42.77%
Best Local Similarity: 29.79%
Query Match: 18.95%
Indels: 122
Gaps: 19
DB:
```

US-09-985-689A-2 (1-434) X US-08-894-818B-4 (1-1566)

```
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 24 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGGTTCTGGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACAATAGGAATAATTGACACTGGAAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCAAGGAAAAGTA-----ATTGGGTGGGTAGATTGTTCAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGGAGTTATCCATACGATGACCATGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 230
QY 78 yAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAAGTAATGCAAGTACAAAGGAAATGGCTCCAGGAGCTTAAGCTGGCGGG 290
QY 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAGGTTCTAGTGCCGATGTTCTTGGAAAGCATATCTACTATAATTAAAGGAGTTGA 350
QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGGCGCTTGTATAACAAGATAAGTACGGAATTAAGTCAATTAATCTTCTCTGCTTC 410
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 411 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGAGCGGGATGC 470
QY 140 rAspSerArgAsnValAlaAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla 160
Db 471 T-----GGATTAGTTGTTGGTTGGCGC 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 495 TGGAAACAGTGGACCTCAACAAGTATACAATCGTTCTCCAGCAGCTGCAACCAAGTTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAGTATGA 581
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATTACAAGCTTCTCAAGCAGAGGCGCAACTGCAGACGCGAGCTTAAGCCTGAGGT 641
QY 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
```

```
Db 642 TGTGCTCCAGAACTGGATTAATGTCCAGAGCAAGT-----GGAACCTAGCAT 692
Qy 240 eTriAlaasnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
Db 693 GGGTCAACCAATTAATGACTATTACACAGAGCTCTGGACATCAATGACCACTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgLysPheValLysAsnArgGlyLeThrPr 280
Db 753 CGTAGCTGGTATTGACAGCCCTCTGCTCCAA-----GCACACCC 791
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAla----- 293
Db 792 GAGTGGACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTATATCGTAAAGCC 851
Qy 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 852 AGATGAATAGCCGATATAGCTACCGGTGCA-----GGTAGGGT 890
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
Db 891 TAATGATACAGGCTATAAAC-----TAGATAACTATGCAAGCTAGTGTTCCTGG 944
Qy 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTCCCAACAAAGGACGCAAACTCACCAGTTGTTATTAGCGAGCTTCGTTCTG 1004
Qy 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATATATCTGGGCAATGCCAAT----- 1035
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTACCCTACGATCCCAATGGAAACCAAG----- 1077
Qy 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1078 -----GTTGACTACTCTTACACCCCTACTATGATTCGAAAGGTGTTGTTATTA 1127
Qy 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCAACTGATGAAATGACATGACATTAAGTTGATAGCTACAGC-----GGAAG 1178
Qy 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTCAAGTAGATGTGTAAGT 1206
```

RESULT 9

```
US-08-894-818B-34
; Sequence 34, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1962 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-34
```

```
Alignment Scores:
Pred. No.: 1,32e-34 Length: 1962
Score: 42550 Matches: 140
Percent Similarity: 42.77% Conservative: 61
Best Local Similarity: 29.79% Mismatches: 147
Query Match: 18.95% Indels: 122
DB: 3 Gaps: 19
```

US-09-985-689A-2 (1-434) x US-08-894-818B-34 (1-1962)

```
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyCl 24
Db 420 GTCTGAGCTCAAGTTATGCGCACTTACGTTGGAACTGGGATATGATGGTCTCGGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 53
Db 522 AGATCTCCAAAGGAAAAGTA-----ATTGGTGGGTAGATTTTGTCAATCG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATACGATGACCATGACATGGAATCATGCTTCAATAGCAGCTGG 626
Qy 78 YAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGTCCAGGAGCTAAGCTGGCGGG 686
Qy 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGTTCTAGTGGCCGATGGTCTCTGGAACCATATCTACTATATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCGGTTGATAACAAAGATAAGTACGGAATTAAGTCAATTAATCTTCTCTGTTTC 806
Qy 132 a-----AlaValAsnGlyValaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGGTACTACGCTCTAAGCTAGGCTGTTTAATGCGAGCTGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla 160
Db 867 T-----GGATTAGTTGTTGGTTCGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
```

891 TGAACAGTGGACCTAACAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
951 TACAGTTGGAGCC-----GTTGCAAGTATGA 977
200 nHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
978 TGTATTAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGGCTTAAGCTGAGGT 1037
220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
1038 TGTGTCTCCAGGAACCTGATATATGCTCCAGAGCAAGT-----GGAACTAGCAT 1088
240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
1089 GSGTCACCAATTAATGACTATTACACAGCAGCTCTGGACATCAATGGCACTCTCA 1148
260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
1149 CGTAGCTGTATTGACGCCCTCTTGCTCCAA-----GCACACCC 1187
280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
1188 GAGCTGGACTCCAGACAAAGTAAACACGCCCTCTAGAACTGCTGATATCGTAAAGCC 1247
294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
1248 AGATGAATAGCCGATATAGCTTACCGTACGTCGA-----GGTAGGGT 1286
310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
1287 TAATGATACAGAGCTATAAC-----TACGATAACTATGCAAGCTAGTCTTCACTGG 1340
330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProbe 345
1341 ATATGTTGCCAACAAAGGCGCAAACTCACAGTTTCGTTATTAGCGGAGCTTCGTTCT 1400
345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
1401 AACTGCCACATTATACGGACATGCCAAT----- 1431
365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
1432 -AGCGACCTTGACTTTACCTCTACGATCCCAATGGAACCCAG----- 1473
385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
1474 -----GTTGACTACTCTTACACCGCCTACTATGGATTCCGAAAGGTTGGTTATTA 1523
405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
1524 CAACCCCACTGATGACATGGACATTAAGTTGTAAGCTACAGC-----GGAAG 1574
425 oGlnAsnPheSerLeuAlaIleValAsn 434
1575 TGCAAACTATACAGTAGATGTGGTAGT 1602

RESULT 10

US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyoza
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15
Alignment Scores:
Pred. No.: 1,32e-34 Length: 1962
Score: 425.50 Matches: 140
Percent Similarity: 42.77% Conservative: 61
Best Local Similarity: 29.79% Mismatches: 147
Query Match: 18.95% Indels: 122
Gaps: 4

US-09-985-689A-2 (1-434) x US-09-445-472-15 (1-1962)

Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlycInGlyGl 24
Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATCATGTTCTCGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db 480 CACATAGGAATATTGACACTGGAATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAAGGAAAGTA-----ATTGGTGGGTAGATTTCATCATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATFACCATGACCATGACATGGAACATCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 YAsnGlyValAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAAGGAAATGGCTCCAGGAGCTAAGCTGGCGGG 686
Qy 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGGTCTTAGTGCCGATGTTCTTGAAGCATATCTACTATATTAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCCTTGATTAACAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTTGGTTC 806
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGCTACTGACGCTTAAGTCAGGCTGTTAATCGACGTTGGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTGTGTGGTTGGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 891 TCGAAACAGTGGACCTCAACAGTATACAAATCGGTTCTCCAGCAGCTCCAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 TACATTTGGAGCC-----GTTGCAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATTAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGTCTCCAGGAACCTGATATATGCTCCAGAGCAAGT-----GGAACCTAGCAT 1088

QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCA 1148
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGGTATGCGAGCCCTTGCTCCAA-----GCACACCC 1187
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGGGACTCCAGACAAATAAAGCAAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgGva 310
Db 1248 AGATGAATAGCGGATATAGCTAGCGTGCA-----GGTAGGTT 1286
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlnSerSerAlaLeuSerThrSe 330
Db 1287 TAATGCATACAAAGGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTTCCTGG 1340
QY 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCCAACAAAGCAGCAACTCACAGTTCGTTATTAGCGGAGCTTCGTTCTGT 1400
QY 345 ulysIleSerLeuValTyrPserAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTATATCGGACAAATGCCAAT----- 1431
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAACACAG----- 1473
QY 385 eSerAlaProPheAspAsnAsnTyrAspGlyArgAsnValGluAsnValPheIleAs 405
Db 1474 -----GTTGACTACTCTTACACCGCCTACTATCGATTGCAAGGTTGTTATTA 1523
QY 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACATGACAAATTAAGGTTGTAAGCTACAGC-----GGAAG 1574
QY 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1575 TCACAACTATCAAGTAGATGTGGTAAGT 1602

RESULT 11

US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B

; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-6

Alignment Scores:

Pred. No.: 1,89e-33 Length: 1977
Score: 414.50 Matches: 134
Percent Similarity: 43.07% Conservative: 68
Best Local Similarity: 28.57% Mismatches: 147
Query Match: 18.46% Indels: 120
DB: 3 Gaps: 19

US-09-985-689A-2 (1-434) x US-08-894-818B-6 (1-1977)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGl 24
Db 420 GTCTGAGCTCAAGTATGCGCACTTACGTTGGAACCTGGATATGATGTTCTGGAAT 479
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db 480 CACATAGGAATAAATGACACTGGAATGCAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAGGAAAGTA-----ATTGGGTGGGTAGATTTGTCAATGG 566
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATACGATGACCATGACATGGAACCTCATGTAGCTTCAATAGCAGCTGG 626
QY 78 yAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAGGGAATGGCTCCAGGAGCTAAGCTGCGGG 686
QY 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGGTTCTAGGTGCGCATGGTTCCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIle----- 125
Db 747 GTGGGCGGTGTGATAACAAAGATAAGTACGGAATTAGGTCATTAAATCTTCTTCTGTTTC 806
QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyValTyrThrTh 140
Db 807 AAGCCAGAGCTCCGACGGAACCGACTCCCTCAGTCAGCGCGCTCAACACGCTGGGACGC 866
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaI 160
Db 867 C-----GGTATAGTAGTCTCGTGGCGCGC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIleAsnAlaI 180
Db 891 CGGCAACAGCGGCGCGCAACACCTACACCTCGCTCGCTCACCCTCGCGCGCGCGAGCAT 950

```
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 AACCGTGGTGCA-----GTTGACAGCAACA 977
QY 200 eHisValAlaGlnPheSerArgGlyProThrIlyAspGlyArgIleIlyProAspVa 220
Db 978 CAACATCGCCAGCTTCTCAGAGGGGACCGCCGCGGAGGCTCAAGCGGAGT 1037
QY 220 iMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 CCGTCCCGCGGGTGGATCATAGCCCGCGCGCCAGC-----GGAACACGAT 1088
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIl 260
Db 1089 GGGACCCCGGATAAAGAGCTACTACCAAGGCGCTCTGGAACCATGGCCACCCCGCA 1148
QY 260 eValAlaGlyIleAsnValAlaGlnLeuArgGluHisPheValIlyAsnAsgIlyIleTh 279
Db 1149 CGTTTCGGGCGTGGCGCGCTCATCTCCAGCCAC-----1185
QY 279 rProIly-----ProSerLeuLeuIlyAlaLeuIleAlaGlyAla-----293
Db 1186 -CCGAGCTGGACCCCGGCAAGAGTGAAGCCGCTCATCGAGACCGCGCATAGTCGC 1244
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyAr 309
Db 1245 CCCAAGGAGATAGCGGATCGCTACGTGG-----GGTAG 1283
QY 309 gValThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSerSerAl 326
Db 1284 GGTGAACGTCTACAGGCCATCAAGTACGACGACTACGCCAGCTCACCTTCCCGGCTC 1343
QY 326 aLeuSerThrSerGlnIlyAlaThrTyrThrPheThrAlaThrAlaGlyLysProLeuIly 346
Db 1344 CGTCGGCGCAAGGAGCGCCACCCACCTTCGACGTCAGCGGCGCACCTTCGTGAC 1403
QY 346 sIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAs 366
Db 1404 CGCCACCTCTACTGGAC-----ACGGGCTCGAG 1433
QY 366 nAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPheSe 386
Db 1434 CGACATCGACCTCTACTCTACGCCCAAGGAGGAGGAG-----1473
QY 386 rAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnSe 406
Db 1474 -----GTTGACTACTCTACACCGCTACTACGCTTCGAGAAGTCTGCTACTACAA 1526
QY 406 rProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyProCl 426
Db 1527 CCCGACCGCGGAACCTGGAGCTCAAGTCTGCTACGCTACAG-----GGCGCGGC 1577
QY 426 nAsnPheSerLeuAlaIleValAsn 434
Db 1578 GAACATACAGGTCTGACGCTGCTAGC 1602

RESULT 12
US-09-000-016-3
; Sequence 3, Application us/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
US-09-000-016-3

Alignment Scores:
Pred. No.: 1.59e-27 Length: 2539
Score: 359.50 Matches: 141
Percent Similarity: 42.07% Conservative: 58
Best Local Similarity: 29.81% Mismatches: 165
Query Match: 16.01% Indels: 109
DB: 3 Gaps: 20

US-09-985-689A-2 (1-434) x US-09-000-016-3 (1-2539)
QY 2 AspValAlaArgGlyIleValIlyAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCGTCGGCAGATCGGCGCCCAAGCGGTGTCGCC-----GGCTACGACGCG 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGCGTGAAGATCGCGCTCTGACACCGGTGTCGACACGAG-----1054
QY 42 MethisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGGCGCGGTGACCGCTCAAGAACTTCACCGCGCGCGCGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
Db 1112 GCCGCGACACAGGTGGGCGCACCGCACCGCTCGATCGCGGCGGCGACGGGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAGGCGCAGTACAGGGGTGCGACCGCGCGCGCTCTCAACGCGCAAGGTC 1231
QY 98 MetAspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTCGGCGAGACTCCCGCATCTCGCGCGCATGGAGTGG 1285
QY 118 AlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
```


Db 1286 GCGCGCGCGAGGCGCGACGTCGTCCACATGAGCCTGGGC-----GGCATG 1333
Qy 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetThr 155
Db 1334 GACACACCGGAGACCGCGGTGGAGCGCGGTGCGACAAGCTGTCCGCGAGAGGCG 1393
Qy 156 IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCTCTGTTGGCATCGCGCGCGCACAGGCGCGCGAG-----TCGATCGTTCGCC 1447
Qy 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGAGCGCGCTCACGTCGCGCC-----1480
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerArgGlyPro---ThrLys 212
Db 1481 -----GTCGACGACAAAGCAAGCTCGCGACTTCTCTCCACGCGCGCGCTCGGC 1534
Qy 213 AspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer 232
Db 1535 GACGCGCCATCAAGCGGACGTACCGCTCCCGGCGTGGACATCACGCGCGCTCGCG 1594
Qy 233 SerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGCGAACGACATCGCGCAGAGGTGCGTGAGGACCGCGCGGTCTACATGACCATCTCC 1654
Qy 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GCGACGTGATGCGACCGCGCACGTGCGGCGCGCGCGCTCTCTCTCTCTCTCTCTCT 1711
Qy 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCCGACTGCACTCGCGCGCACTGAAGCGGCGC 1747
Qy 290 IleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArg 309
Db 1748 CTCACGGCTCCACCAAGGCGCG---AAGTACCCCGTTCGAGCAGGGTTCGGCGCG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSer--- 328
Db 1805 ATCCAGCGCGACAGCGCTCCAGCAGACCGTATCGCGCGCGCGCTCTCTCTCTCTCTCT 1864
Qy 329 -----ThrSerGlnLysAlaThrTyrThr 336
Db 1865 GCGCTCCAGCAGTGGCGCGACACCGACGAGCGGTGTCACCAAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGCACCGAGCGTCACGCTGAAGCTGACGTGACCGCGCACCGACCCCAAGGCG 1984
Qy 354 -----AlaProAlaSer-----ThrThrAlaSerValThrLeuValAsnLeuAsp 369
Db 1985 AAGCGCGCGCGCGCGGTCTTCTACGTCGCGCGCACCGC-----2026
Qy 370 LeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPhe 389
Db 2027 -----GTACCGCTCCGCGCGCGCG-----AGGCGCTCCGTC 2059
Qy 390 AspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSer 409
Db 2060 GACATGACCGCGCACACCGCG-----CTCGCGCGCAGCGTGGAC 2098
Qy 410 GlyThrTyrThrIle-----414
Db 2099 GCGCGGTACTCGGTACGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2158
Qy 415 -----GluValGlnAlaTyrAsnValProVal 423
Db 2159 GCGGTGACGCGCGAGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2197

RESULT 13

US-09-514-340-3

; Sequence 3, Application US/09514340

; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRINDINE D
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
; US-09-514-340-3

Alignment Scores:
Pred. No.: 1.59e-27 Length: 2539
Score: 359.50 Matches: 141
Percent Similarity: 42.07% Conservative: 58
Best Local Similarity: 29.81% Mismatches: 165
Query Match: 16.01% Indels: 109
DB: 4 Gaps: 20

US-09-985-689A-2 (1-434) x US-09-514-340-3 (1-2539)

Qy 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCGCGCGGAGATCGCGCGCCCAAGCGGTGTTCGCC---GGTACGACGCG 1009
Qy 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGCTCTGCGACACCGGTGTGACACGAGC-----1054
Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAGGCGCGGTGACCGCGTCCAGAGACTTCACCGCGCGCGCGC 1111

QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
DB 1112 GCGGCGGACAAAGTGGCGCACCGCACCGCTCGATCGCGGGGCGACCGGGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
DB 1172 CAGTCCAGGCGCAAGTACAGGCGCTCGCACCGCGCGGCGATCCTCAACGCGCAAGTCC 1231
QY 98 MetAspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
DB 1232 CTCGACGACTCGGTT-----TTCGCGACGACTCCGCGATCCTCGCGCGCATCGAGTGG 1285
QY 118 AlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
DB 1286 GCGGCGGCGCGCGCGCGCGCTCGTCCACCATGACCTGGCG-----GGCATG 1333
QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetThr 155
DB 1334 GACACACCGGAGACCGCGCTGGAGCGGCGGTGCAACAAGCTCTCCGCGCGAGAAGGCG 1393
QY 156 IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
DB 1394 GTCTGTTCCTCCATCG 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
DB 1448 GCGAGCGCGCGCGCGCGCGCTCACCGCTCGCGCGCG-----1480
QY 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerArgGlyPro-----ThrIys 212
DB 1481 -----GTGCGAGCAAGGACAAAGCTCGCGCGCTCTCTCCACCGCGCGCGCGCTCGCG 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer 232
DB 1535 GACGGCGCATCAAGCGCGACGTACCGCTCCCGCGGTGGACATCACCGCGCGCTCGCGCG 1594
QY 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
DB 1595 GAGGCGCAACATCGCGCGCGAGGTGCGTGGAGGACCGCGCGCGGTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
DB 1655 GGCACGCTCATGCG 1711
QY 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----289
DB 1712 -----CACCGCGACTGGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1747
QY 290 IleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg 309
DB 1748 CTCACCGGCTCCACCAAGGCGCG-----AAGTACACCGCGTTCGAGCAGGTTTCGGCGCG 1804
QY 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlnSerSerAlaLeuSer---328
DB 1805 ATCCAGGCGCGACAAGGCGCTCCAGCAGACCGGTGATCGCGCGCGCGCGCGCGCGCGCG 1864
QY 329 -----ThrSerGlnLysAlaThrTyrThr 336
DB 1865 GCGGTCCAGCAGTGGCGCGCGCACCGACGCGCGCGTCCACNAGAGCGTACCTACCGC 1924
QY 337 PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----353
DB 1925 AACCTCGGCGACCCAGGACGTCAGCTGAAGCTGACGTGACCGCGCGCGCGCGCGCGCGCG 1984
QY 354 -----AlaProAlaSer-----ThrThrAlaSerValThrLeuValAsnAspLeuAsp 369
DB 1985 AAGCGCGCGCGCGCGCGCTTCTTCACGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2026
QY 370 LeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPhe 389
DB 2027 -----GTGACCGCTCCCGCGCGCGCG-----AGCGCTCCGCTC 2059
QY 390 AspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSer 409

DB 2060 GACATACCGCGCACCGCG-----CTCGGCGCGCGGTGGAC 2098
QY 410 GlyThrTyrThrIle-----414
DB 2099 GCGCGCTACTCGCGCTACGTGTGTCGCCAGCGGCGCGCGCGTCCGCGACGCGCGCC 2158
QY 415 -----GluValGlnAlaTyrAsnValProVal 423
DB 2159 GCGGTCCAGCGCGAGTCTGAGTCTGACGACGTGACCGTC 2197
RESULT 14
US-09-000-016-1
Sequence 1, Application US/09000016
Patent No. 6143541
GENERAL INFORMATION:
APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERI
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
STRAIN: A-914
ORGANISM: Streptomyces viridosporus
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores: 6.23e-27 Length: 2809
Pred. No.:

Score: 354.50 Matches: 139
Percent Similarity: 41.49% Conservative: 56
Best Local Similarity: 29.57% Mismatches: 173
Query Match: 15.79% Indels: 103
DB: 3 Gaps: 17

US-09-985-689A-2 (1-434) x US-09-000-016-1 (1-2809)

QY 2 AspValAlaArgGlyLeuValValAlaAspValAlaGlnSerSerTyGlyLeuTyArgly 21
DB 953 GACAGCTCGTGGGCGAGATCGGCGCCGCCAAGCGTGTTCGGCC---GGCTACGACGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
DB 1010 AAGGGCGTGAAGATCGCGCTCTGGACACCGGTGTGGACAGCAGC----- 1054
QY 42 MetHisGluAlaPheArgGlyLeuIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsn 61
DB 1055 ---CATCGGACCTGAAGGGCGGTGACCGCGTCCAGAACTTCCACCGCGCGCCCGGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
DB 1112 GCCGCGCAAGGTGGGCGACGCGACCGACCGATCGCTCGATCGCGCGGCGACGGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
DB 1172 CAGTCCAAAGGCAAGTACAGGGGTGCGCACCGCGCGCGGATCCTCAACGGCAAGGTC 1231
QY 98 MetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
DB 1232 CTCGACGACTCCGGT-----TTGCGGACGACCTCCCGCATCTCCGCGCGCATCGAGTGG 1285
QY 118 AlaPheSerAlaGlyAlaArgile-HisThrAsnSerTrpGlyAlaAlaValAlaGlyAl 137
DB 1286 GCGCGCGCGAGGCGCGCGTCTCAACATGAGCTGGCGCGCATGGACACACCGGAG 1345
QY 137 aTyThrThrAspSerArgAsnValAspAspTyThrValArgLysAsnAspMetThrIleLe 157
DB 1346 ACCGCGCGTGGAGGCGCGG-GTCGACAACTGTCCGCGAGAGGGCGTCTGTTCGC 1404
QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
DB 1405 CATCGCGCGCGGCAACGAGGCGCGGAG-----TCGATCGGTTCGCGCGCGCGGGA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAs 197
DB 1459 CGCGCGCGCTCACCGTCCGCGC-----GTCGA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216
DB 1486 CGACAGGACAGCTCGCGCGATCTCTCTCCACCGCGCCCGCGCTCGCGGCGCGCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrTyThrIleLeuSerAlaArgSerSerLeuAlaPr 236
DB 1546 CAAGCGCGACGTACCGCTCCCGCGGTGGACATCACGCGCGCTCCGCGAGGCAACGA 1605
QY 236 cAspSerSerPheTrpAlaAsnHisAspSerLysTyAlaTyMetGlyThrSerMe 256
DB 1606 CATCGCGCGAGGTCGTCGAGGACCGCGCGGTATCATGACCATCTCCGCGACGTGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
DB 1666 GCGCAGCCCGCGCTCGCGCGCGCGCGCGCTCTCTGAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----lleAlaGlyAl 293
DB 1712 -----CACCCCGACTGACCTCCGCGCGAACTGAAGGCGCGGTCTCACCGGCTC 1758
QY 293 aAlaAspValGlyLeuGlyTyProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
DB 1759 CACCAGGGCGGC---AGTACACCCGTTGAGCAGGTTTCGGCGCGGATCCAGGCGGA 1815
QY 313 pLysSerLeuAsnValAlaTyValAsnGluSerSerAlaLeuSer----- 328

DB 1816 CAAGCGCTCCAGCAGACCGGTGATCGCGACCGCGGTCTCGGTGAGCTTCGGCGTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyThrPheThrAlaTh 340
DB 1876 GTGGCGGCACACCGCAGCAGCGGTACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----AlaPr 355
DB 1936 CCAGGACGCTACGCTGAAGCTACGTCGACCGCCACCGACCCCAAGGGCAGAGCGCGCC 1995
QY 355 oAlaSer-----ThrThrAlaSerValThrLeuValAsnAspLeuLeuValIleTh 373
DB 1996 GCGCGGCTTCTTCACGCTGGCGGCCACACG-----GTGAC 2031
QY 373 rAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTr 393
DB 2032 CGTCCCGCGGGCGGCG-----AGCGCTCCGTCGACATGACCGC 2070
QY 393 pAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyTh 413
DB 2071 CGACACCGCG-----CTCGGCGCAGCGTGGACGGCGGTACTC 2109
QY 413 rIle----- 414
DB 2110 GCGGTACGTGTGCGCACGCGGCGGCGGCGAGCGGTCCGACGCGCGCGTGCAGCG 2169
QY 415 -GluValGlnAlaTyAsnValProVal 423
DB 2170 CGAGGTGAGTCTGACGACGTGACCGTC 2197

RESULT 15

US-09-514-340-1
Sequence 1, Application US/09514340
Patent No. 6361987
GENERAL INFORMATION:
APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE D
ITS EXPRESSION PRODUCT

NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Bonack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; STRAIN: <Unknown>
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID No: 1:
US-09-514-340-1

Alignment Scores:
Pred. No.: 6,23e-27 Length: 2809
Score: 354.50 Matches: 139
Percent Similarity: 41.4% Conservative: 56
Best Local Similarity: 29.57% Mismatches: 173
Query Match: 15.79% Indels: 103
DB: 4 Gaps: 17

US-09-985-689A-2 (1-434) x US-09-514-340-1 (1-2809)

QY 2 AspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
DB 953 GACACGTCTCGCGGAGATCGCGCCGCCCAAGGCGTGGTCGGCC---GGCTACGACGGC 1009

QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
DB 1010 ARGGGGTGAAGATCGCGCTCTGGACACCGGTGTGACACGAGC----- 1054

QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
DB 1055 ---CATCCGACCTGAAGGCGCGGTGTGACCGGTCAAGAACTTCCACCGCGCGCGCGC 1111

QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
DB 1112 GCGGGCGAAGTGGGCGCACGGCACCCACGTCGCTCGATCGCGGGGACAGGGCGCC 1171

QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
DB 1172 CAGTCCAAAGGGCAAGTACAAAGGCGTGCACCGCGCGCGCGATCTCAACGCGCAAGGTC 1231

QY 98 MetAspSerSerGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
DB 1232 CTCGACGACTCCGT-----TTCGGCGACGACTCCGGCATCTCGCGGCGATGAGTGG 1285

QY 118 AlaPheSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
DB 1286 GCGGGCGCGGCGGCGCGCGACGTCGTCAACATGAGCTGGGCGCATGGACACACCGGAG 1345

QY 137 aTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleIle 157
DB 1346 ACCGACCGGTGGAGGCGCG-GTCACACAGTGTCCGCGGAGAGGGCGTCTGTTCGC 1404

QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
DB 1405 CATCGCGCGCGCAACGAGGCGCCGAG-----TCGATCGTTCGCCCGGACGCGCGA 1458

QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
DB 1459 CGCGCGCCCTCACCCTCGGCGCC-----GTGGA 1485

QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgIle 216
DB 1486 CGACAGGCAAGCTCGCGACTTCTCTCCACCGGCCCGCGCTCGGCGGCGGCCAT 1545

QY 216 eLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaPr 236
DB 1546 CAAGCGGAGGTCAACCGCTCCCGCGGTGGACATCACGCGCGCTCGCGGCGGCAACGA 1605
```

```

QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
DB 1606 CATCGGCCAGGAGGTCTGGTGGAGGCGCGCGCTACATGACCATCTCCGGCACGTGAT 1665

QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
DB 1666 GCGCACCCCGCACGTGCGCGGCGCGCGCCCTCTCTGAAGACGACGAG----- 1711

QY 276 ggIylIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
DB 1712 -----CACCCCGACTGGACCTCCGCCGAAGTGAAGGGCGCGCTCACGGCTC 1758

QY 293 aAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
DB 1759 CACCAAGGGCGCGC---AAGTACACCCCGTTCGAGCAGGGTTCGGCGCGGATCCAGGCCGA 1815

QY 313 pLysSerLeuAsnValAlaTyrValAsnGlnSerSerAlaLeuSer----- 328
DB 1816 CAAGGCGCTCCAGCAGACCGTGTATCGCCGACCGGCTCTCGGTGAGCTTCGGGCTCCAGCA 1875

QY 329 -----ThrSerGlnLysAlaThrTyrThrPheThrAlaTh 340
DB 1876 GTGGCGCGCACACCGACGAGCGCGTCCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935

QY 340 xAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----AlaPr 355
DB 1936 CAAGGAGCTCAGCTGAGCTGACGTGACCGCCACCGACCCCAAGGGCAAGCGGCGCC 1995

QY 355 oAlaSer-----ThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleTh 373
DB 1996 GCGGGGCTTCTTCACGCTGGCGCGCCACCG-----GTGAC 2031

QY 373 xAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTr 393
DB 2032 GTCCCGCGCGGCGCGC-----ACGCGCTCCGTGACATGACCGC 2070

QY 393 pAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrTh 413
DB 2071 CGACACCGCG-----CTCGGCGGACGCGTGGACGCGCGTACTC 2109

QY 413 rIle----- 414
DB 2110 GGCGTACGTGTGCGCACGCGCGCGGCGGAGACGCTCCGACGCGCGCGCGTGCAGCG 2169

QY 415 -GluValGlnAlaTyrAsnValProVal 423
DB 2170 CGAGGTGAGTCTGACGACGTGACCGTGC 2197
```

Search completed: April 4, 2004, 12:01:36
Job time : 92.5977 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 328.004 seconds
(without alignments)
4948.852 Million cell updates/sec

Title: US-09-985-689A-2

Perfect score: 2245

Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVVPGQNFSLAIVN 434

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09985689/runat_31032004_161809_4271/app.query.fasta_1.3498
-DB=Published Applications NA -OPFM=fastap -SUFFIX=tnpb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 UNITS=bits -SPAP=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09985689 @CGN_1_1_601 @runat_31032004_161809_4271
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubnpa/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubnpa/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubnpa/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubnpa/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubnpa/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubnpa/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubnpa/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubnpa/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubnpa/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubnpa/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubnpa/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubnpa/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubnpa/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubnpa/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubnpa/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubnpa/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubnpa/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

ALIGNMENTS

RESULT 1

US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIEFUMI
; APPLICANT: SAKKI, KAISUHIISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938US0
; CURRENT APPLICATION NUMBER: US/10/385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

1	2191	97.6	1305	15	US-10-385-662-1	Sequence 1, Appli
2	457.5	20.4	1977	13	US-10-090-624-11	Sequence 11, Appli
3	425.5	19.0	1236	13	US-10-090-624-2	Sequence 2, Appli
4	425.5	19.0	1962	13	US-10-090-624-15	Sequence 15, Appli
5	360.5	16.1	3624	14	US-10-156-761-5701	Sequence 5701, Ap
6	360.5	16.1	9025608	14	US-10-156-761-1	Sequence 1, Appli
7	345.5	15.4	3417	14	US-10-156-761-3306	Sequence 3306, Ap
8	345.5	15.4	9025608	14	US-10-156-761-1	Sequence 1, Appli
9	338.5	15.3	135838	14	US-10-314-657-1	Sequence 1, Appli
10	308.5	13.7	1329	9	US-09-974-300-1934	Sequence 1934, Ap
11	308.5	13.7	1560	15	US-10-084-846A-113	Sequence 113, App
12	308.5	13.7	59816	15	US-10-084-846A-1	Sequence 1, Appli
13	308.5	13.7	59816	15	US-10-084-846A-2	Sequence 2, Appli
14	307	13.7	4765	15	US-10-090-624-5	Sequence 5, Appli
15	283	12.6	3788	10	US-09-927-827-33	Sequence 33, Appl
16	277	12.3	3743	10	US-09-927-827-39	Sequence 29, Appl
17	277	12.3	2192	12	US-10-424-599-112429	Sequence 112429,
18	273	12.2	1306	9	US-09-966-921A-1	Sequence 1, Appli
19	273	12.2	1330	9	US-09-966-921A-5	Sequence 5, Appli
20	270	12.0	3303	14	US-10-156-761-5384	Sequence 5384, Ap
21	261	11.6	1971	9	US-09-974-300-1935	Sequence 1935, Ap
22	256	11.4	840	14	US-10-209-812-1	Sequence 1, Appli
23	254	11.3	2166	12	US-10-344-231-17	Sequence 17, Appl
24	254	11.3	2166	12	US-10-363-332A-17	Sequence 17, Appl
25	253	11.3	1485	9	US-09-974-300-1938	Sequence 1938, Ap
26	246	11.0	1140	8	US-08-322-678-11	Sequence 11, Appl
27	246	11.0	1140	8	US-08-322-678-12	Sequence 12, Appl
28	246	11.0	1140	16	US-10-323-324-11	Sequence 11, Appl
29	246	11.0	1140	16	US-10-323-324-12	Sequence 12, Appl
30	246	11.0	1143	14	US-10-313-853-6	Sequence 6, Appli
31	246	11.0	2588	12	US-10-344-231-20	Sequence 20, Appl
32	246	11.0	2588	12	US-10-363-332A-20	Sequence 20, Appl
33	240.5	10.7	3884	10	US-09-927-827-34	Sequence 34, Appl
34	236.5	10.5	1497	8	US-08-322-678-6	Sequence 6, Appli
35	236.5	10.5	1497	9	US-09-060-854B-1	Sequence 1, Appli
36	236.5	10.5	1497	14	US-10-033-325-1	Sequence 1, Appli
37	236.5	10.5	1497	15	US-10-228-572-1	Sequence 1, Appli
38	236.5	10.5	1497	15	US-10-423-649-1	Sequence 1, Appli
39	236.5	10.5	1497	16	US-10-323-324-6	Sequence 6, Appli
40	234.5	10.4	1140	15	US-10-146-905A-9	Sequence 9, Appli
41	234.5	10.4	1149	15	US-10-146-905A-7	Sequence 7, Appli
42	234.5	10.4	1494	14	US-10-104-693-1	Sequence 1, Appli
43	233	10.4	2267	12	US-10-344-231-18	Sequence 18, Appl
44	233	10.4	2267	12	US-10-344-231-18	Sequence 18, Appl
45	232.5	10.4	4338	9	US-09-891-711-3	Sequence 3, Appli

```
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Bacillus sp. KSM-KP43
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 1,34e-235 Length: 1305
Score: 2191.00 Matches: 419
Percent Similarity: 99.54% Conservativity: 13
Best Local Similarity: 96.54% Mismatches: 2
Query Match: 97.59% Indels: 0
DB: 15 Gaps: 0

US-09-985-689A-2 (1-434) x US-10-385-662-1 (1-1305)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 1 AATGATGTTGGCGTGGATTGTCAAAGCGGATGGCTCAGACGAGCTACGGTTGTAT 60
QY 21 GlyGlnGlyGlnLeuValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 61 GGAACAAGGACAGATCGATCGGTTGCCGATACAGGCGTTGATACAGGTGGCAATGACAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysLeuThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 121 TCGATCGATGAAGCTTCGCGGGAATAATCTGCAATATATGCTTGGACGACGAT 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 181 AATGCCAATGATACGAATGCTATGTGTACGATGTGGCTGGCTCGCTATTAGAAACGCG 240
QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 241 TCCACTTANTAAAGAAATGGCGCTCAGCGGAATCTAGTCTTCAATCATCATGATAGC 300
QY 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
DB 301 GGTGGGGGACTTGGAGGACTTACCTCGAATCTGCAAACTTATTCAGCCCAAGCATACT 360
QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
DB 361 GCTGGTGCCAGAAATTCATCAAACTCTGGGGAGCAGAGTGAATGGGCTTACACAA 420
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 421 GATTCACAGAAATGTGGATGACTATGTGGCGCAAAATGATATGACGATCTTTTCGCTGC 480
QY 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaIle 180
DB 481 GGGAAATCAAGACCGGACCGGAAATCATAGTGCACCGACGACGCTAAATGAATA 540
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 541 ACAGTCGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACATAATCAAC 600
QY 201 HisValAlaGlnPheSerSerArgGlyProThrIleAspGlyArgIleLysProAspVal 220
DB 601 CATGTGGCAGATTCTCTTACGTGGACCGCAAGAGATGGAGGATCAAAACCGGATGTC 660
QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
DB 661 ATGCAACGGGAAGCTTCATCTATACGCAAGATCTCTCTTGCACCGGATCTCTCTTC 720
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
```

```
DB 721 TGGCGAACCATGACAGTAATAATATGATACATGGGTGGAACTCCATGGCTACACCGATC 780
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 781 GTTGTCTGGAACGTCGACAGCTTCGTGAGCATTTTGTGAAACACAGAGGCATCACACCA 840
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
DB 841 AAGCCTTCTCTATTAAAGCGGACCTGATGGCGGTGACGCTGACATCGGCTTGGCTAC 900
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 901 CCGAACGGTAACCAAGGATGGGACGAGTGACATTGATAAATCCCTGAACGTTGCCTAT 960
QY 321 ValAsnGlnSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
DB 961 GTCAACGAGTCCAGTTCTCTATCCACAGCCAAAAGACGACTCTGTTTACTTGTCTACT 1020
QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
DB 1021 GCGGCAAGCCTTTGAAATCTCCCTGGTATGCTCTGATGCCCTGCGAGCACAACTGCT 1080
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
DB 1081 TCCGTACGCTTGTCAATGATCTGGACCTTGTCATTAACCGCTCCAAATGGCACAGTAT 1140
QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTyrAspGlyArgAsnValGlu 400
DB 1141 GTAGGAATGATCTTACTTCTGCGCATACATGATACTGGATGGCGGCAATACGTAGAA 1200
QY 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1201 AATGATATTATTAATGACCAACAAAGCGGACGCTATACAAATTGAGGTACAGGCTTATA 1260
QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
DB 1261 GTACCGTTGGACACAGACTTCTGTTGGCAATTGTGNAT 1302
```

RESULT 2

```
US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11
```

```
Alignment Scores:
Pred. No.: 6.75e-41 Length: 1977
Score: 457.50 Matches: 136
Percent Similarity: 44.66% Conservativity: 69
Best Local Similarity: 29.63% Mismatches: 153
Query Match: 20.38% Indels: 101
DB: 13 Gaps: 17
```


Db 171 TAGGAGTATCCATGACATGACATGGAATCTAGCTTCAATAGCAGCTGG 230
Qy 78 yAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAGGGAATGGCTCAGAGCTAAGCTGGCGG 290
Qy 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAAGTTCTAGGTGGCGATGGTCTGGAGACATACTACTATAATTAAGGAGTTGA 350
Qy 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGGCGGTGTGATAACAAAGATAAGTACGGAATTAAGTCATTAATCTTTCTTGGTTC 410
Qy 132 a-----AlaValAsnGlyAlaValThrTh 140
Db 411 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTATATGACGCTGGGATGC 470
Qy 140 xAspSerArgAsnValAspTyrValArgIleAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTTGTTGTGGTTCGCGC 494
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla1 180
Db 495 TGGAAACAGTGGACCTTAAAGATATACANTCGTCTCCAGCAGCTGCAGCAAGTTAT 554
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAGTATGA 581
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATTAAACAGCTTCTCAGCAGAGGGCCAACTGCAGACGGCAGCTTAAACCTGAGGT 641
Qy 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGTCTCCAGGAACCTGGATAATTGTCGCAGCAAGT-----GGAACATAGCAT 692
Qy 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro1 260
Db 693 GGGTCACCAATTAATAGCTATTACACAGCAGCTCTCGACATCAATGGCACTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGTATTGCAGCCCTCTTGTCCAA-----GCACACC 791
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGACCTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAGCC 851
Qy 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 852 AGATGAATAGCCGATATAGCTACGTGCA-----GGTAGGT 890
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
Db 891 TAATGCATACAGGCTATAAC-----TAGCATAACTATGCAAGCTAGTGTCTACTCG 944
Qy 330 xGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCCMACAAAGCAGCCAACTCACCAGTTGTTATTAGCGGAGCTTCGTTCGT 1004
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATTATCTAGTGGCAATGCCAAT----- 1035
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1036 -AGCACCTTGATCTTACCTCTACGATCCCAATGGAAACAG----- 1077
Qy 385 eSerAlaProPheAspAsnAntTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405

Db 1078 -----GTTGACTACTCTTACACCCCTACTATGATTGGAAGAGTTGGTTATTA 1127
Qy 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCAACTGATGGACATGGACAATTAAAGTTCTTAAGCTACAGC-----GGAAG 1178
Qy 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGGTAAGT 1206
RESULT 4
US-10-090-624-15
; Sequence 15, Application US/10090624
; Publication No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-15
Alignment Scores:
Pred. No.: 2,59e-37 Length: 1962
Score: 425.50 Matches: 140
Percent Similarity: 42.77% Conservative: 61
Best Local Similarity: 29.79% Mismatches: 147
Query Match: 18.95% Indels: 122
DB: 13 Gaps: 19
US-09-985-689a-2 (1-434) x US-10-090-624-15 (1-1962)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTGACGCTCAAGTTATGGCAACTTACGTTGGAACTGGGATATGATGGTTCTGGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACAATAGGAATAATTGACACTGGCAATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAGGAAAGTA-----ATTGGGTGGGTAGATTGTTCATGG 565
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 yAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAGGGAATGGCTCCAGAGCTAAGCTAGCGCGG 686
Qy 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAAGTTCTAGTGGCGGATGGTCTTGGAGCATATCTACTATATAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132

```
Db 747 GTGGCGCTGTGATATACAAAGATAAGTACGGAATTAAAGTCTAATCTTTCTCTGTGTC 806
Qy 132 a
Db 807 AAGCCAGAGCTCAGATGCTACTGACGCTCTAAGTCAGGCTGTAAATGACGGTGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrLeuPheAlaAl 160
Db 867 T-----GGATTAGTGTGTGGTGGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
Db 891 TGGAAACAGTGGACCTTCAACAGTACAAATCGGTTCTCCAGCAGCTGCAACCAAGTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATTAAACAGCTTCTCAGCAGAGGCGCACTGCACAGCGCAGGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTCTCTCAGGAAACTGGATAAATCTGCCAGAGCAAGT-----GGAACCTAGCAT 1088
Qy 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI 260
Db 1089 GGGTCAACCAATTAACTACTATTACACAGCAGCTCTGGGACATCAATGGCAATCCTCA 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGTATTGACGCGCTCTTCTCCAA-----GCACACCC 1187
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1189 GAGCTGAGCTCAGCAAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAGCC 1247
Qy 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 1248 AGATGAAATAGCCGATATAGCTACGGTGCA-----GGTAGGGT 1286
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
Db 1287 TAATGCATACAGGCTATAAAC-----TACGATAACTATGCAAGAGTAGTGTTCACCTGG 1340
Qy 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAGGCAGCCAACTCACCAGTTCTGTTATTAGCGGAGCTTCTGTCGT 1400
Qy 345 uLysIleSerLeuValTrpSerAspAlaProIleSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTATATCTGGACAATGCCAAT----- 1431
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAAACCAAG----- 1473
Qy 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1474 -----GTTGACTACTCTTACACCGCTACTATGATTCGAAAGAGTGTGTTATTA 1523
Qy 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCACTGATGGAACTAGACAAATTAAGTTTGAAGTACAGC-----GGAAG 1574
Qy 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGTGAAGT 1602
```

RESULT 5

US-10-156-761-5701

```
; Sequence 5701, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701

Alignment Scores:
Pred. No.: 1,26e-29 Length: 3624
Score: 360.50 Matches: 129
Percent Similarity: 44.39% Conservative: 53
Best Local Similarity: 31.46% Mismatches: 161
Query Match: 16.06% Indels: 67
DB: 17 Gaps: 17

US-09-985-689a-2 (1-434) x US-10-156-761-5701 (1-3624)
```

```
Qy 8 VallyAlaAspValAlaGlnSerSer-----Tyr 17
Db 556 GTCGAGCGCACATGTCGCGGACGACAAACGCGCAGATCGGTACGCGGCGCCGCGTGGGACGCC 615
Qy 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 616 GGGCTACGGCGACGCGGCTCACCCTGCGCGTGCTGCACACGCGGCTGCACACC----- 669
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 670 -----ACTCACCCCGACCTCGCGCGCGGGTGTCCCGGACGACGAGCTTCATC 717
Qy 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 718 GACGGGAGGAGTCCGCCACCGCAACGCGCACGGACCCACGTCACCTCGACCTCGCGC 777
Qy 78 GlyAsnGlyAla-----ThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 93
Db 778 GGCAGCGCGCGCCCTCCGACGCGCAGCGGCGCGGCGTCCGCGCGTGCACGCTCGCC 837
Qy 94 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThr 113
Db 838 GTCGGCAAGTGTCTCAGCGACCGACCGCGCG-----GGAAGCAGATCCCGATC 885
Qy 114 LeuPheSerGlnAlaPheSerAla-----GlyAlaArgIleHisThrAsnSerTrp 130
Db 886 ATCGCGGGCATGGAATGGCGCGCGCGGACGTCGCGGACGATCGTCTCGATGAGCCTC 945
Qy 131 GlyAla-----AlaValAsnGlyAlaTyrThrThrAsp-----SerArgAsnValAsp 146
Db 946 GGATCGACCGAGCGCAGCGG-----ACCGACCCCATGCCCGCGCGCTCGAC 996
Qy 147 AspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsn 166
Db 997 ACCCTCTCCGAGGAGACCGCGCGCTCTCTGTCGCGCGGGAACACCGCGTGC----- 1053
```


QY 266 AlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrProLysProSerLeuLeu 285
 |||||
 Db 6919557 GCGCTCTCGCGACAGCACACCGCGCTGACCGCGCGCGCTCAAGACGCGCTGATG 6919616
 |||||
 QY 286 LysAlaAla-----LeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGly 303
 |||||
 Db 6919617 TCCAGCTCCGAGCAACTCGACGCCCTCCGTATATCATGTTGGGGGCGGTTCGGTCAAGTGTG 6919676
 |||||
 QY 304 AsnGlnGlyTyrGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsn 322
 |||||
 Db 6919677 CCGGACGCGGTGGCGCGCGGTACCGCGACGCGCGCGACCTCGGTTCACCGC 6919736
 |||||
 QY 323 GluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThrAlaGly 342
 |||||
 Db 6919737 TGGCCCCCATAGCGCGCATCGACCGCTCACAGACGCGTCACTCACTCAACTCTCCGAC 6919796
 |||||
 QY 343 LysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerVal 362
 |||||
 Db 6919797 ACGAGGTGAGTGGAGCTCGCCGTGGGGGCGCGCGGTGTC-----GCC 6919847
 |||||
 QY 363 ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro-----AsnGlyThr 378
 |||||
 Db 6919848 ACCCTCGGC-----GACACCGCACTCACCGTGGCGCGCGCCACGCGCGCGCCACC 6919898
 |||||
 QY 379 ArgTyrValGlyAsnAspPheSerAlaPro 388
 |||||
 Db 6919899 ACCGTGACCGCGACGGCTCCAAGGCTCG 6919928
 |||||

RESULT 7

US-10-156-761-3306
 ; Sequence 3306, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 3306
 ; LENGTH: 3417
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(3417)
 US-10-156-761-3306

Alignment Scores:

Pred. No.: 5,57e-28 Length: 3417
 Score: 345.50 Matches: 130
 Percent Similarity: 43.68% Conservative: 53
 Best Local Similarity: 31.03% Mismatches: 158
 Query Match: 15.39% Indels: 80
 DB: 14 Gaps: 13

US-09-985-689A-2 (1-434) x US-10-156-761-3306 (1-3417)

QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrClyArg 37
 |||||
 Db 745 GGGTACGACGCGAAGGGCGTCAAGATCGCCGCTCGGACACCGGTGTCGAC----- 795
 |||||

QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeudgly 57
 |||||

Db 796 -----GCGACCCACCCGCGACCTCAAGGACCAAGTGGCGGAGTCCAAAGAACTTCTCC 846
 |||||
 QY 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 |||||
 Db 847 GCGCGCGCGACGCGCCGCCACCATTCGTGTACGGCAGCAGCTCGCTTCATCGCGGG 906
 |||||
 QY 78 GlyAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
 |||||
 Db 907 GGCACCGCGCGCAAGTCCAAACGCAAGTACAAGGTGTGCGCGCGCGCGCGATCCCTC 966
 |||||
 QY 94 PheGlnSerIleMetAspSerSerGlyLeuGlyLeuGlyGlyLeuProSerAsnLeuGlnThr 113
 |||||
 Db 967 AACGGCAAGGTCTCCGACGACACCGGC-----TCGGCGAGACTTCGGCATCTTGGCC 1020
 |||||
 QY 114 LeuPheSerGlnAlaPheSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAl 133
 |||||
 Db 1021 GGCATGGAGTGGCGCGCGCGAGCGCGCGCTGTCAACCTGAGCGCTGGCGCGCGC 1080
 |||||
 QY 133 aValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAs 153
 |||||
 Db 1081 GACACCCCGAGATCGACCGCTGGAGCGGAG-GTCAACAAGCTTCTCCGAGGAGAGGG 1139
 |||||
 QY 153 pMetThrIleLeuPheAlaAlaGlyAsnGlyProAsnGly---GlyThrIleSerAl 172
 |||||
 Db 1140 CATCTCTTCGGCATCGCGCGCGCAACGAGGCGAGTTCGGCGAGCAGCATCGGCTC 1199
 |||||
 QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
 |||||
 Db 1200 CCGGGCAGCGCGCGCGCTACCGTCCGTCGCGC----- 1236
 |||||
 QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrIly 212
 |||||
 Db 1237 -----GTGACGACGACGACGAGCTGGGTCTTCTCCAGCGCGCGCGCGCT 1286
 |||||
 QY 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgse 232
 |||||
 Db 1287 GGACGCGCGCATCAAGCGCGACGTACCACCGCGCTGGAGTTCACCGCGCGC----- 1341
 |||||
 QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLys-----TyrAl 249
 |||||
 Db 1342 -----GCGCGCGCGCGCGCGTATCCAGGAGGTGGCGCGAGACCGCGAGTACT 1397
 |||||
 QY 249 aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr 269
 |||||
 Db 1398 CACCATCTCGGTACGTGCGATGCGACCGCGCATGTGCGGGCGCGCGCGCATCTCAA 1457
 |||||
 QY 269 gLiuHisPheValLysAsnArgGlyIleThrProLysProSerLeu-----LeuLysAl 287
 |||||
 Db 1458 CGACGACGAC-----CCCAACTGCTGCTTCGCGAGTCAAGGG 1496
 |||||
 QY 287 aAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr 307
 |||||
 Db 1497 CGCGCTGACCGGCTCCGCGAAG-----GGCGCAAGTACACGCGCTTCAGCAGGGCTC 1550
 |||||
 QY 307 pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlySerSerAlaLe 327
 |||||
 Db 1551 GGGCCGTATCGCGCTCGACAGCGCATCAAGCAGTCCGTGATCGCAACCCCGAACTCGGT 1610
 |||||
 QY 327 uSerThrSerGlnLysAlaThrTyrPhe-ThrAlaThrAlaGlyLysProLeuLysI 347
 |||||
 Db 1611 GAGC-----TTGCGCATCCAGCAGTGGCGCGCACACCGACGA 1646
 |||||
 QY 347 leSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA 367
 |||||
 Db 1647 CAAGCGCGTCAACCCAGCAGTCACTACCGCACTTCGCGCAGCGAGTGCATC-ACGCTGA 1705
 |||||
 QY 367 spLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerA 387
 |||||
 Db 1706 ACTTCGGTCCGCGCGCGACCAACCCCAAGGGCGTC-----GCCG 1744
 |||||
 QY 387 laProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnSerP 407
 |||||

RESULT 9
US-10-314-657-1
; Sequence 1, Application US/10314657
; Publication No. US2003017588A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; TITLE OF INVENTION: Synthases and Methods of Use
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22

```
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

Alignment Scores:
Pred. No.: 258e-25 Length: 135638
Score: 343.00 Matches: 128
Percent Similarity: 43.06% Conservative: 58
Best Local Similarity: 29.63% Mismatches: 144
Query Match: 15.28% Indels: 102
DB: 14 Gaps: 17

US-09-985-689A-2 (1-434) x US-10-314-657-1 (1-135638)
Qy 6 GlyLeuValValAlaAspValAlaGlnSer----- 15
Db 8067 GGCAGGGTGAAGGCGGATCTCGCGGACTCCACCGCCAGATCGCGCGGAGAGGTATGG 8126

Qy 16 SerTyGlyLeuTyGlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThr 35
Db 8127 GCGAGGGCGACACCGCGGAGAGCGTGAAGTTCGATGCTCGACAGCGGCGGACACC 8186

Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysile-----ThrAlaLeu 53
Db 8187 -----GAACACCGGACCTGCTGGCGAGGTGTCGACAGCGCGCAGC 8228

Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAla 73
Db 8229 TTCGTCCCGCGGAGGACGATCGCC---GACTACACGGCCACCGGACGCGATCGCC 8285

Qy 74 GlySerValLeuGlyAsnGlyAlaThrAsn-----LysGlyMetAlaProGln 89
Db 8286 TCGACCATCTCGGACGCGGACGCGCCCTCCGACGCGAAGAGCGGGTGTGCTCCGCG 8345

Qy 90 AlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGly----- 103
Db 8346 GCCCGCGCTGCTCGCGGAAGGTGCTCAACTCCGAGGCGACGCCAGGATCGTGGATC 8405

Qy 104 LeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAla 123
Db 8406 ATCGCGGCGCATGGAG-----TGGGCGCGCGCGGACCAAGGCC 8444

Qy 124 ArgIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaThrThrAsp----- 141
Db 8445 AGGATCATCAGCATGAGCTCGGC-----GGCGCGGTGACAAAGACGACCCGATG 8495

Qy 142 SerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 161
Db 8496 AGCCAGGCGGTGCGAAGTCACTAGCCACACGACGCGCGGTGTTCGTGATCGCGCGGC 8555

Qy 162 AsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 181
Db 8556 AACGCGCGCCGAC-----TCCATCAGACGCGCGGTGCGGACGACTCGCGCTGACC 8609

Qy 182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHis 201
Db 8610 GTCGCGCGCC-----GTCGATCTCCACCGACAGC 8636

Qy 202 ValAlaGlnPheSerSerArgGlyProThrLys---AspGlyArgIleLysProAspVal 220
Db 8637 CTCGCGGACTTCTCCAGCGAGGCGCGGTGACGCGGAGCGGGGCTGAAGCGGAGATC 8696

Qy 221 MetAlaProGlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 8697 ACCGCGCGCGGCTGCACATCGTCGCGCGCGCTCGCACTAAGCGCGGCTCCGCGGTAC 8756

Qy 241 TrpAlaAsnHisAspSerLysTyAlaTyMetGlyThrSerMetAlaThrProIle 260
Db 8757 -----TACACCATGATGCGGACGCGCATGCGGACGCGGAC 8795

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 8796 GTCGCGGGTTCGCGCGGCTCTCCGCGCGAGACCCCGAGCTGGACGCGGACCGAGCTC 8855

Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTy 300
Db 8856 AAGGAGGCACTGTCAGCAGCGCCCAAGCAAGCGCGCG-----TAC 8897

Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTy 320
Db 8898 ACCCGGTACCGCGCGGCGCGCGGTCTCGACGCGCGCGCGCGGTGCACACCGAGTTC 8957

Qy 321 ValAsnGlnLysSerAlaLeuSer-----ThrSer 330
Db 8958 TTCGCCACACGACGCGCTACTCCGCTTCACACGTTGCGCCCGCGAGCGCGGAGACC 9017

Qy 331 GlnLysAlaThrTyThrPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
Db 9018 GATGTCGCGAGCGTGACGTACACCAACGTCGCGGACGCGCGGTCAAGCTCAACCTG 9077

Qy 351 TrpSer-----AspAlaPro 355
Db 9078 GTCACGCGCACCGTCCCGCGCGGTGTTCACGCTCTCCGAGGACCATGTACCGTGCC 9137

Qy 356 AlaSerThrAlaSerValThrIleuValAsnAspLeuAsp----- 369
Db 9138 GCGCAGCGCACCGCCACGCTCACTGACCGCGGTCTGCAACAGTGGCGGCGGACGAG 9197

Qy 370 -----LeuValIleThrAlaProAsnGlyThr 378
Db 9198 TCGGTGACGCGCGTGTATCACCGTACGAGCGGACG 9233

RESULT 10
US-09-974-300-1934
; Sequence 1934, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Alignment Scores:
Pred. No.: 83e-28 Length: 1329
Score: 338.50 Matches: 105
Percent Similarity: 43.53% Conservative: 43
Best Local Similarity: 30.88% Mismatches: 109
Query Match: 15.08% Indels: 83
DB: 9 Gaps: 13

US-09-985-689A-2 (1-434) x US-09-974-300-1934 (1-1329)
Qy 11 AspValAlaGlnSerSerTyGlyLeuTyGlyGlnIleValAlaValAlaAsp 30
Db 406 GAATGCTCAGAACAAATCAGCGCTGACAGCAAGAGGAGTGCAGTCTGCTATTGAT 465
```

31 ThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArgGlyLysile 50
466 ACAGCGGTA-----TACCCTCAGAGAGATCTTGAAGGACGAGATC 504
51 ThrAlaLeuTyr---AlaLeuGlyArgThrAsnAlaAsnAlaAspThrAsnGlyHisGly 69
505 AGGCGCTTTTCAAGACTTTTCAACCCAGAGAACAGAACCTATGATGACAAATGGGACCGC 564
70 ThrHisValAlaGlySerValLeuGlyAsnGlyAlaAlaThrAsn-----LysGly 85
565 ACACACTGCGCGGTGATCTTGGGAACCGGAGCGGCTCATCGGTCAGTACCGCGGA 624
86 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 105
625 CCGTCTCTCTGAGCAGCACTTGTGCGTGTAAAGTATTGGACAAAATGGGATCC---GGA 681
106 GlyLeuProSerAsnLeuGln-----SerTyrGlyAlaAlaValAsnGly 136
682 TCGCTCGAAACGCTCAATTCAAGCGTAGATTGGTGCATTCATTCATTAAGGAATCCT 741
113 -----ThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsn-128
742 GATGATCCGATCGACATTATTCAATGTCATTGGGTGCAGAGCCCTTGGCTACGAGAT 801
129 -----SerTyrGlyAlaAlaValAsnGly 136
802 GAAGAAAGACATCCAGTCGTAAAGCTGTTTCAGCAGATGGACCGCAGCAT---855
137 AlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrile 156
856 -----GTTGTA 861
157 LeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAla 176
862 TGTGCGGACCGCGCACTCCGCTCTGTATGCCAAACGATTCGACGCGGCTGTGCAC 921
177 LysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAla 196
922 AGCAAGATTATTACAGTCGAGCTTGGATGAC---AGGATACACTCAGCGGAGGAT 978
197 AspAsnIleAsnHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIle 216
979 GAGCAT-----GTCGCTCTTATTCAAGCAGAGGCGGACAAATCTATGTCAGATC 1029
217 LysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer-----SerLeu 234
1030 AACCGGACTGTGTTGACCGGCACAAATATTACGTCGCTGTTCCCGGATCTTT 1089
235 AlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 254
1090 CTCGATAAGCTGCAAAAAACAAACAGAGTCGCGCACAAATATATGACATTGTCGGAACC 1149
255 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 274
1150 TCGATGCTACGCGCATCTGGCAGGAATTCGCGCA-----CTTATCCTT 1194
275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 294
1195 CAGCAAGCCCGGCGCACAGAACTGTGTAAGTCAAAACAGCTGTCTAATGGACGCT---ACC 1251
295 AspVal-----GlyLeuGlyTyrProAsnGlyAsnGln 305
1252 GATTATTGGAAGATCGCGATCCAAATGTTTACGTCAGGGTACATCAACGACGACAA 1311

RESULT 11

US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUELENWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS

TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
FILE REFERENCE: 1974-005
CURRENT APPLICATION NUMBER: US/10/084,846A
CURRENT FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: PCT/EP01/09815
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: DE 101 09 166.4
PRIOR FILING DATE: 2001-02-25
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn Ver. 3.2
SEQ ID NO 113
LENGTH: 1560
TYPE: DNA
ORGANISM: Streptomyces viridochromogenes
FEATURE:
OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:

Pred. No.: 2,44e-24 Length: 1560
Score: 308.50 Matches: 115
Percent Similarity: 39.34% Conservative: 51
Best Local Similarity: 27.25% Mismatches: 159
Query Match: 13.74% Indels: 97
DB: 15 Gaps: 17

US-09-985-689A-2 (1-434) x US-10-084-846A-113 (1-1560)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 451 GGGCAGGAGTGCAGCGGCTACGTCATCGACACCGCGCTC-----CGC 492
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
DB 493 ATCACCACACGCACTTCGGCGCGCGGCTCTCTACGGCTACGACGCCATCGACCAACGAC 552
60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
553 AACACCGCCAGGACGCGCCAGCGGCAGCGCAGCGCAGCGCTGCGCGGACGCTCGCGGCAAC 612
QY 80 GlyAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
DB 613 GCC-----TACGGGTGCGCAAGAGCCAGATCGTAGGCTGCCGCTGCTGAAC 663
QY 100 SerSerGly-----GlyLeuGlyGlyLeuProSerAsn 110
DB 664 AACTCGCGCCAGGACCCACCGCCAGGTGTCGCGGCATCGACTGGTCCGCGGGAAC 723
QY 111 LeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTyr 130
DB 724 -----GCCGTCAAGCGCGCGCTCGCAACATGTCCTC 756
QY 131 GlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArg 150
DB 757 GCGCGCGCGCGCACAGCGCTCGACACAGCGCGCTACGCAAC-----GCCATG 804
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIle 170
DB 805 GCCTCCGCGCCACTTCG 864
QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
DB 865 TCA---CCGCGACCGGTCAACGAGGCCATCACGCTCGCGCGCGCGCGCGCGCGCGCGCG 921
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
DB 922 AAGCGCGGTACTCCAACTACGCTCCGCTC-----954
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrIleLeu 228
DB 955 -----GACCTCTTCGCGCGCGCGGTTCGTCATCACC 984


```
QY 229 SerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyr 248
D 985 TCGGCTC-----TGAACCTCAAGCGACTCGCGGACC 1014
QY 249 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
D 1015 AACACCATCTCCGGTACGTGATGCGACCCCGCACGTCGGGGGCGCGCGCTC--- 1071
QY 269 ArgGluHisPheValValAsnArgGlyThrProLys-----ProSerLeuLeu 285
D 1072 -----CACCTCGCGCGCAACCCCTCGGCGACCCCGTCCACGTCGCGCACGCGCTACG 1125
QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGln 305
D 1126 TCCGCGCGCACACCGCGCTGCTCACCACCCCGGACGGGCTCGCCCAAC----- 1176
QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu----- 323
D 1177 -----CGGCTCCTGTACGTGCGCGCGGCGCAC 1203
QY 324 -----SerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAla 339
D 1204 GACCACTCCCGCGCGCGCTTCGAGAACCCGCTGACTACACATCAGCGACAACTCC 1263
QY 340 ThrAlaGlyLysProLeuLysLysSerLeuValTrpSerAspAlaProAlaSerThrThr 359
D 1264 ACGTTCGAGTCCCGGTGAGCTCTCCGCGTCTCCGGCAACGCGCCCTCGGCCCTCGCC 1323
QY 360 AlaSerVal-----ThrLeuValAsnAspLeuLeuValIleThrAlaPro 375
D 1324 GTAGAGGTCCACATGTCACACGATACATCGCGGACCTCCAGTCCAGTCCAGTCCGCCCC 1383
QY 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGly 395
D 1384 GACGGACCGGTACACGCTCAAGTCGACGCGACC-----GGCGGCGAGTTCGGACAA 1437
QY 396 ArgAsnAsnValGluAsnValPheLeuAsnSer-----ProGlnSerGlyThrTyrThrIle 414
D 1438 ATCAACACCACTGCTCGGTGAACGCTCTCCGAGGCGGCGCAACGGCACGCTGGAACCTG 1497
QY 415 GluVal 416
D 1498 CGGGTG 1503
```

RESULT 12

```
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-1
```

Alignment Scores:

```
Prod. No.: 5,64e-22 Length: 59816
Score: 308.50 Matches: 115
Percent Similarity: 39.34% Conservative: 51
```

```
Best Local Similarity: 27.25% Mismatches: 159
Query Match: 13.74% Indels: 97
DB: 15 Gaps: 17
US-09-985-689A-2 (1-434) x US-10-084-846A-1 (1-59816)
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
D 56648 GGCACGGGAGTGACGGCGTACGTATCGACACCGGGCTC-----CGC 56689
QY 41 SerVetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
D 56690 ATCACCCACAGCGACTTCGGCGCGCGGCTCTCTACGGTACGACGCCATCGACACGAC 56749
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
D 56750 AACACCGCCGACGCGCACCGCCACGCGACGTCGCGCGGACGTCGCGCGGCAAC 56809
QY 80 GlyAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
D 56810 GCC-----TACGGCGTCGCCAAGAACGACATCGTAGGCGTCGCGTGTGAAC 56860
QY 100 SerSerGly-----GlyLeuGlyGlyLeuProSerAsn 110
D 56861 AACTCGGCGGACGCGCACCGCCAGTCGTCGCGGATCGACTGGGTGCGCCGGAAC 56920
QY 111 LeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
D 56921 -----GCCGTCAAGCGCGCTCGCCCAACATGTCCCTC 56953
QY 131 GlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspTyrValArg 150
D 56954 GCGCGCGCGCGCGCACCGCCCTCGACACGCGCGTACGCAAC-----GCCATG 57001
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
D 57002 GCCTCGCGGTCACTTCGCGGTGGCGCGGACGACGATCGCACACCTCCACGAGG 57061
QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
D 57062 TCA---CCCGCACGCGTCACCGAGGCGCATCACGTCGCGCGGACGACGCTCGCGCGCC 57118
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
D 57119 AGCGCGGTACTCCAACTACGCTCGTCTC----- 57151
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeu 228
D 57152 -----GACCTCTCGCCCCCGTTCGTCCATCAAC 57181
QY 229 SerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyr 248
D 57182 TCGGCC-----TGGAACTCAAGCGACTCGCGGACC 57211
QY 249 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
D 57212 AACACCATCTCCGGTACGTGATGGGACACCCCGCACGTCGGCGGCGCGCGGCTC--- 57268
QY 269 ArgGluHisPheValLysAsnArgGlyThrProLys-----ProSerLeuLeu 285
D 57269 -----CACCTCGCGCGCAACCCCTCGGCGCACCCGCTCCAGTCCGCGCGCTGACG 57322
QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGln 305
D 57323 TCCGCGCGCACCGCGGCTCGTCACCAACCCCGCGCGGCTCGCCCAAC----- 57373
QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu----- 323
D 57374 -----CGGCTCCTGTACGTGCGCGCGGCGCAC 57400
QY 324 -----SerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAla 339
D 57401 GACCACTCCCGCGCGCGCTTCGAGAACACCGGTCGACTACACGATCAGCGACCACTCC 57460
```

QY 340 ThrAlaGlyLysProLeuLysSerLeuValTrpSerAspAlaProAlaSerThrThr 359
 DB 57461 ACGGTGAGTCCCGGTGAGTCTCCGGCGTCTCCGCAACGGCCCTCGCCCTGCC 57520
 QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 DB 57521 GTAGAGTCCATCGTCCACAGTATCATCGCGACCTCCAGTCCAGTCCAGTCC 57580
 QY 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAsnAsnTrpAspGly 395
 DB 57581 GACGGCAGCGGTACAGCTCAAGTGTACGGCACC-----GGCGGCGAGTTCGGACAAC 57634
 QY 396 ArgAsnAsnValGluAsnValPheIleAsnSer---ProGlnSerGlyThrThrIle 414
 DB 57635 ATCAACACCACTGCTCGGTGAACGCTCTCGAGGGCGGCCAACGGCAGCTGGAAC 57694
 QY 415 GluVal 416
 DB 57695 CGGGTG 57700

RESULT 13

US-10-084-846A-2/c
 ; Sequence 2, Application US/10084846A
 ; Publication No. US2004006026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: WEITNAUER, GABRIELE
 ; APPLICANT: MUHLNUEG, AGNES
 ; APPLICANT: TREFFER, AXEL
 ; APPLICANT: BECHTHOLD, ANDREAS
 ; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
 ; FILE REFERENCE: 1974-005
 ; CURRENT APPLICATION NUMBER: US/10/084,846A
 ; PRIOR FILING DATE: 2003-02-25
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4
 ; PRIOR FILING DATE: 2001-02-25
 ; NUMBER OF SEQ ID NOS: 120
 ; SOFTWARE: Patent in Ver. 3.2
 ; SEQ ID NO 2
 ; LENGTH: 59816
 ; TYPE: DNA
 ; ORGANISM: Streptomyces viridochromogenes
 US-10-084-846A-2

Alignment Scores:

Pred. No.: 5,64e-22 Length: 59816
 Score: 308.50 Matches: 115
 Percent Similarity: 39.34% Conservative: 51
 Best Local Similarity: 27.25% Mismatches: 159
 Query Match: 13.74% Indels: 97
 DB: 15 Gaps: 17

US-09-985-689A-2 (1-434) x US-10-084-846A-2 (1-59816)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 3169 GGGCAGGAGTACGGCGGTACGTATCGACACCGCGTC-----CGC 3128
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
 DB 3127 ATCAACCAAGCGCTTCCGGCGCGCGCTCTTACGGCTACGCGCATCGACACAGCAG 3068
 QY 60 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 DB 3067 AACACCGCCAGGAGCGCCACGGCAGCGACGCGTGGCGGCGAGTTCGGCGCAAC 3008
 QY 80 GlyAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
 DB 3007 GCC-----TACGGCGTCCGCAAGAAGCAAGATCGTAGGCGTCCGCGTGTGAAC 2957
 QY 100 SerSerGly-----GlyLeuGlyGlyLeuProSerAsn 110

RESULT 14
 US-10-090-624-5

DB 2956 AACTCCGGCCAGGCGACACCGCCAGGTCGTCCGGCATCGACTGGGTCCCGCGAAC 2897
 QY 111 LeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
 DB 2896 -----GCCGTCAAGCGCGCGTCCGCCAACATGTCTCCCTC 2864
 QY 131 GlyAlaAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArg 150
 DB 2863 GGGGGCGCGCCACACCGCCCTCGACACGGCGGTACGCAAC-----GCCATG 2816
 QY 151 LysAsnAspMetThrIleLeuPheAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
 DB 2815 GCCTCCGGCTCACTCCGTCGCGTGGCGCGCGCACAGTCCGACCAACGCCCTCCAGAG 2756
 QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
 DB 2755 TCA---CCGACCGCGTCACCGAGGCGCATCGGTCCGCGCGACCGACGTCGGACGCC 2699
 QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
 DB 2698 AAGCGCGCTACTCCAACTACGCTCCGCTCCTC----- 2666
 QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeu 228
 DB 2665 -----GACCTCTTCCGCCCGCGGTTCGTCCATCACC 2636
 QY 229 SerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyr 248
 DB 2635 TCGGCC-----TGAACCTCAAGCGGACTCGCGAC 2606
 QY 249 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
 DB 2605 AACACCATCTCCGTGTCGTGTCGACACCGCGTGGCGCGCGCGCGCTC--- 2549
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
 DB 2548 -----CACCTCGCGCGCAACCCCTCGCCACCGCTCCAGTCCGACGGCGCTGACG 2495
 QY 286 LysAlaAlaLeuIleAlaGlyAlaAspValGlyLeuGlyTyrProAsnGlyAsnGln 305
 DB 2494 TCCCGCGCCACACCGCGCTGTCCACCAACCGCGCGCGCGCTCGCCCAAC----- 2444
 QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu----- 323
 DB 2443 -----CGGCTCTCTAGTCTCGCGCGCGCGCAC 2417
 QY 324 -----SerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAla 339
 DB 2416 GACCACCTCCGGCGCGCGCTTCGAGACACCGGTGACTACAGTACGACGACCACTCC 2357
 QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThr 359
 DB 2356 ACGGTGAGTCCCGGTGACGCTCTCCGGCGTCTCCGCAACCGCGCTCGGCCCTCGCC 2297
 QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 DB 2296 GTAGAGTTCACATCGTCCACACGTATCATCGCGACCTCCAGTCCAGTCTCGCCCCC 2237
 QY 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAsnAsnTrpAspGly 395
 DB 2236 GACGCGCGGTACACGCTCAAGTGTACGGCACC-----GGCGGAGTTCGGACAAC 2183
 QY 396 ArgAsnAsnValGluAsnValPheIleAsnSer---ProGlnSerGlyThrTyrIle 414
 DB 2182 ATCAACACCACTGCTCGGTGAACGCTCTCTCGAGGGCGGCCAACGGCAGCTGGAAC 2123
 QY 415 GluVal 416
 DB 2122 CGGGTG 2117

RESULT 15
US-09-927

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2279.84 seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-2
Perfect score: 2245
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVOAYNVFVGQNFSLAIVN 434

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09985689/runat.31032004.161807.4168/app.query.fasta.1.3498
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=ext -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689@cgn.1.1.7084/brunat.31032004.161807.4168 -NCPU=6 -ICPU=3
-NO_MMALP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hc.*
9: gb_est.*
10: gb_est2.*
11: gb_hc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gssl.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	249	11.1	1605	13	BQ622771	BQ622771 CC Contig
2	231.5	10.3	4198	11	AK029048	AK029048 Mus muscu
3	230	10.2	640	12	BJ395336	BJ395336 BJ395336
4	219.5	9.8	594	12	BJ393752	BJ393752 BJ393752
5	216.5	9.6	532	29	TA319G10P	TA319G10P T. brucei
6	208.5	9.3	2141	13	BQ142519	BQ142519 Contig M
7	203	9.0	771	14	CA320325	CA320325 UI-M-FW0-
8	200	8.9	508	28	AQ652212	AQ652212 Sheared D
9	199.5	8.9	601	12	BJ387574	BJ387574 BJ387574
10	195	8.7	633	12	BJ369190	BJ369190 BJ369190
11	195	8.7	718	12	BI750157	BI750157 FG02 10G0
12	193.5	8.6	1002	29	CNS06D6B	AL393417 T3 end of
13	188	8.4	565	13	BQ770462	BQ770462 UI-M-F10-
14	186	8.3	716	28	BZ893395	BZ893395 HL2_0177
15	185	8.2	675	14	CF727824	CF727824 UI-M-HB0-
16	183.5	8.2	574	29	TA315H10P	AL490202 T. brucei
17	183.5	8.2	614	9	AJ273402	AJ273402 AJ273402
18	183	8.2	641	12	BJ393925	BJ393925 BJ393925
19	178.5	8.0	650	9	AJ274038	AJ274038 AJ274038
20	178	7.9	530	29	CNS010PO	AL153820 Anopheles
21	178	7.9	895	13	BQ216158	BQ216158 AGENCOURT
22	178	7.9	2121	28	BZ424995	BZ424995 100023066
23	178	7.9	3091	11	BC011275	BC011275 Mus muscu
24	177.5	7.9	681	14	CB690041	CB690041 CEST-54-B
25	177	7.9	576	14	CD295943	CD295943 StrPu691.
26	177	7.9	1572	11	AX107161	AX107161 Zea mays
27	176	7.8	4662	11	BC060627	BC060627 Mus muscu
28	175.5	7.8	712	14	CD311344	CD311344 StrPu691.
29	173.5	7.7	580	9	AJ273745	AJ273745 AJ273745
30	173.5	7.7	583	9	AJ273947	AJ273947 AJ273947
31	173.5	7.7	593	9	AJ273918	AJ273918 AJ273918
32	173.5	7.7	601	9	AJ273921	AJ273921 AJ273921
33	173.5	7.7	616	14	CF138007	CF138007 UI-HF-BNO
34	171.5	7.6	601	9	AJ273050	AJ273050 AJ273050
35	171.5	7.6	604	9	AJ273185	AJ273185 AJ273185
36	171	7.6	449	12	BJ359939	BJ359939 BJ359939
37	170.5	7.6	573	14	CA937626	CA937626 sav42b10.
38	170	7.6	794	14	CF737198	CF737198 UI-M-HD0-
39	169.5	7.6	609	9	AJ272712	AJ272712 AJ272712
40	169	7.5	781	13	BU612128	BU612128 UI-M-EM0-
41	168.5	7.5	545	13	BU575479	BU575479 TgESTzyb8
42	168.5	7.5	610	9	AJ274218	AJ274218 AJ274218
43	168.5	7.5	930	13	EX390734	EX390734 EX390734
44	168	7.5	539	14	CD306363	CD306363 StrPu691.
45	168	7.5	831	14	CA510555	CA510555 UI-R-FU0-

ALIGNMENTS

RESULT 1
BQ622771
LOCUS
DEFINITION CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.
VERSION BQ622771.1 GI:21649940
KEYWORDS EST.
SOURCE Conidiobolus coronatus (Delacroixia coronata)
ORGANISM Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)

AUTHORS Freimoser, F.M., Screen, S., Hu, G. and St. Leger, R.J.
TITLE EST analysis of genes expressed by the zygomycete pathogen *Conidiobolus coronatus* during optimized secretion of proteins
JOURNAL Unpublished (2002)
COMMENT Contact: Freimoser F. M.
 Department of Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA
 Tel: 301 405 16 13
 Fax: 301 314 92 90
 Email: ff34@mail.umd.edu.

FEATURES Location/Qualifiers
 1..1605
 /organism="Conidiobolus coronatus"
 /mol_type="mRNA"
 /strain="ARSEF 512"
 /db_xref="taxon:34488"
 /clone_lib="Conidiobolus coronatus ARSEF 512"
 /notes="Vector: Unizap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and peptone for 18 hours. A cDNA library was constructed in the unidirectional Lambda vector UniZap."

source

ORIGIN

Alignment Scores:
 Pred. No.: 4,05e-14 Length: 1605
 Score: 249.00 Matches: 86
 Percent Similarity: 43.99% Conservative: 42
 Best Local Similarity: 29.55% Mismatches: 83
 Query Match: 11.09% Indels: 80
 DB: 13 Gaps: 14

US-09-985-689A-2 (1-434) x BQ622771 (1-1605)

Qy 6 GlylValValAlaAspValAlaGlnSerSer-----TyrGlyLeu----- 19
 Db 718 GGTGTTCTAAGCTATTGCCGTCACCAAGCAATGCTCTGGGGTCTCTCGTGTGT 777
 Qy 20 -----TyrGlyGlnGly 23
 Db 778 CAACGTGCTAAGCTCGATCTGCTTACTCTTACACCACTGCTGCTCAAGT 837
 Qy 24 GlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHis 43
 Db 838 GTCACGTGTTTTCGTTTAGTACTGCTGTCATGTCGTCAGCCACCAATGAC----- 885
 Qy 44 GluAlaPheArgGlyGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAla--- 62
 Db 886 -----TTCGTGTCGTGTCACCT-----TGGGTACTACACTGCTGCT 924
 Qy 63 -----AsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 925 GGTAGCAACACTGATGCTCAGGTCACGGTCACGTCACCTGCTGCTGCTATTGCTGCT--- 981
 Qy 80 GlyAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet--- 98
 Db 982 -----ACCACCTATGTTGTCGCAAGAGCTAACATTGTCGCTTAAGGTCCTTAGGT 1035
 Qy 99 ---AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuInThrLeuPheSerGln 117
 Db 1036 GATCATGCTCCGATCATACTCTCGAATTATCTCCGTTATCTGCTGGGTGTGAAGCAC 1095
 Qy 118 AlaPheSerAlaGlyAlaArgGlyIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
 Db 1096 -----TCTGTGTCACAAAGTATTCTCTATGATTAGGAGTGTAAGACGATGCT 1149
 Qy 138 TyrThrThrAspSerArgAsnValAspAspThrValArgLysAsnAspMetThrIleLeu 157
 Db 1150 CTTAACACT-----GCTGTTTAAACACGCTGTCAGCAAGGAGTGTGCACTGTTGTC 1200
 Qy 158 PheAlaAlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
 Db 1201 ---GCTGCTGTAACGATAACAGAGATGCTTGTGGATGACTCT---CCCGTCTCTGCTCT 1254

Qy 178 AsnAlaIleThrValGlyAlaThrGlu-----AsnLeuArgProSerPheGlySerTyr 195
 Db 1255 TCCGCCATTACCGTTGGTCCCACTGATCAATGATATAAAGGCTTCACTTCTTAACTTC 1314
 Qy 196 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyGlyProThrLysAspGlyArg 215
 Db 1315 GGTAGTGTGTC----- 1326
 Qy 216 IleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerLeuAla 235
 Db 1327 -----GATATCTTAGCTCTGCTGTCACCAATCTCTCCACC----- 1362
 Qy 236 ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSer 255
 Db 1363 -----TGAAGGATCTAACCATGCCACCAACCACTCTCTGTTACTCT 1407
 Qy 256 MetAlaThrProIleValAlaGlyAsnValAla 266
 Db 1408 ATGGCTGCCCTCACATGCTGCTGTTAGTGTCT 1440
RESULT 2
 AK029048 4198 bp mRNA linear HTC 18-SEP-2003
 LOCUS Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:4732484M11 product:membrane-bound transcription factor protease, site 1, full insert sequence.
 ACCESSION AK029048
 VERSION AK029048.1 GI:26325017
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636
REFERENCE
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159
REFERENCE
 3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, X., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, I., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 2050913
 11076851
REFERENCE
 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 403, 695-690 (2001)
REFERENCE
 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
REFERENCE
 6 (bases 1 to 4198)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayata, N., Hiramoto, K., Hiraoka, I., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsumura, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sojabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Takeda, Y., Tanaka, Y., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

TITLE
JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@sc.riken.go.jp].
URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

source

1. 4198
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4732484M11"
/db_xref="MG1:2391080"
/db_xref="taxon:10090"
/clone="4732484M11"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"

CDS

417..3575
/note="unnamed protein product; membrane-bound transcription factor protease, site 1 (MGD|GI:1927235, GB|NM_019709, evidence: BLASTN, 99%, match=3782) putative"

/codon_start=1
/protein_id="BAC26263.1"

/db_xref="GI:26325018"
/translation="MKLVSTLVLLVLLCKRHLGRLGTRALEKAPCSGHLTLK VEFSTVVEYEVYVAFNGYFTAKARNFISALKSEVENWRIIPRNPSSDPSDFE PIQIKKOKAGLLLEDHPNKRVTPOKRVFSLKFAENPIVPCNETRWSQKWSR PLKASLSLGSGLGHLLEHRRSLRAIPROVAQTLQADVMQMGYTGANRVAVFD TGLSEKHPHFKNVKEWTNWRLLDGLGHGTFVAGVASMRECOGAFDAELHIFR VFTNQVSYTSWFLDAFNAILKMDVLNLSIGGPDNDHPFDVXKVELFANVIVUS AIGNDPGLTLNPDADQMGVIGGIDFEDNIARFSRGMTWELPFGYGRVKPDIV TYGAGVRSGVKNKGRALSGTSVASPVVAGATLLVSTVQKRELVPASVKQALIASA RPLPVNNEFGHGKGLDLRLAYQILSSYKQASLSYIDTECPYMWPCSYIYVG GMPTIVNVTIINGMGVTRIVDKPEWRPFLPQNGDNI EVAFSYSLWPMWSGYLAISI WDVNKAASWEGIAQCHIMTVASPAETELHSGAHTSKLPIKVIITPTPRSKRVL SDVHNLVYPGYGRNRMKNDPLDWDGPHVTNFRDMYHLRSGYEVVLGAPP TCFDATQGTLLLDVSEBEPPEIAKLRRDVLNGLSLVPSDWNTSVMRKVFDE TRQWMPEDTQGNIPALNELLVSNMNGFSDGLYEGEFVLNHDHMYVASGSAKFPF DGVTITQTFKQGLLEVKQETAVVNVPIGLYQIPSEGGRIIVLYGDSNCLDSDSRQ KDCFWLLDALQYTSYGVTPPSLSHSGNRQPPSGAGLAPPERMEGNHLHRSKVL EA HLGDPPKPLPACPKHLSWAKPQPLNTEAPSNLWKHQLSLDLDKVLNPNFRNRPDV RPLSPGESGANDIPGIMPGRYNCEVGVTIPVFAFLGAMVALAFVQVQISKAKSRPKR RPRAKRPPOLAAQHPARTPSV"

polyA_signal

4180..4185

/note="putative"

4198

polyA_site

/note="putative"

ORIGIN

1767 CG-GAGACTTCCTGGGGTCAACATGTCGAGCAAGTCATGG-----1807

Alignment Scores:

Pred. No.:	1,328-11	Length:	4198
Score:	231.50	Matches:	122
Percent Similarity:	39.14%	Conservative:	51
Best Local Similarity:	27.60%	Mismatches:	165
Query Match:	10.31%	Indels:	105
DB:	11	Gaps:	22

US-09-985-689a-2 (1-434) x AK029048 (1-4198)

QY	8	VallysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnLeuValAla	27
DB	1005	CTGCAGGCAGATGTCTGTGG---CAGATGGGATACACAGGCTCTAATGTCAGAGTTGCT	1061
QY	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47
DB	1062	GTTTGTGATCTGGGCTC-----AGTGAGACATCCGCGATTTTAAG	1103
QY	48	GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-----	63
DB	1104	AAT-----GTGAAGGAGAGAGAACCAACTGGACCAATGAGCGGACCTGTG	1145
QY	64	---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThr	82
DB	1146	GATGATGGCTAGGCATGGCACATCTGTTGCGGT---GTGATGCCAGCATGAGGAG	1202
QY	83	AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSer	101
DB	1203	TGCCAAGGATTTGCTCCAGATGCAGATGCACATCTTCAGGGTCTTTACCAACATCAG	1262
QY	102	GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla	121
DB	1263	-----GTGCTTACATCTTGTGTTCTGGATGCCCTTCACTAT	1301
QY	122	GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaThrThrAsp	141
DB	1302	GCCATCCTAAAGAGATGACGCTTCTCAACCTTAGCATCGTGGGCGCCGATTCATGAT	1361
QY	142	SerArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAla	159
DB	1362	CATCCGTTTGTGACAAGGTGTGGGAATTAACAGTACATGTAATGTTGTTCTGCT	1421
QY	160	AlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAla	179
DB	1422	ATTGGCAATGATGACCTCTCTATGGCACTCTGTAATAACCTGCTGATGATGATGATG	1481
QY	180	IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspSerIle	199
DB	1482	ATTGGAGTCGGTGGCATTCAC-----TTTGAAGATTAACATC	1517
QY	200	AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp	213
DB	1518	-----GCTCGCTTTCTCCAGGGGAATGACTACCTGGGAATTACCAAGAGGCTAT	1568
QY	214	GlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSer	233
DB	1569	GGTCGTGTGAAGCTGACATTCCTC-----ACCTATGGTGGTGGAGTGGCGGGTTC	1619
QY	234	LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly	253
DB	1620	GGTGTGAAGGGGGCTGC-----CGTGCACTCTCAGGG	1652
QY	254	ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal	273
DB	1653	ACCAGTGTGCTTCCCGAGTGTGCTGGGGCGCTCACCTTGTAGTAGACAGATACAG	1712
QY	274	LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla	293
DB	1713	AAGCGGAGCTGGTG-----AATCTGCACTGTGAAGCAAGCTTTGATAGCTCAGCC	1766
QY	294	AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlyTyrTrpGlyArgValThrLeuAsp	313


```

/sex="mat A"
/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SP"

ORIGIN
Alignment Scores:
Pred. No.: 7,47e-12 Length: 594
Score: 219.50 Matches: 56
Percent Similarity: 51.3% Conservative: 39
Best Local Similarity: 30.2% Mismatches: 61
Query Match: 9.78% Indels: 29
DB: 12 Gaps: 8

US-09-985-689A-2 (1-434) x BU393752 (1-594)
QY 49 LysileThrAlaLeuTyAlaLeuGlyArgThrAsnAlaAsnAspThrAsnGlyHis 68
DB 21 AAGTTGTAACTATATATACCAATCAACAAGCGAGTAGTAGTAAAGTGGTGTAC 80
QY 69 GlyThrHisValAlaGlySerValLeuGly 81
DB 81 GGTACACATATTTGGTTCTGCACAGGTACTCCAGAGGATTTCTCAGTTAATATTC 140
QY 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 101
DB 141 TCATTAGTGTCTTGCCTGCACTGATGCAAGATTTGCATTTCTTGAATTTGGCAAGTGTCA 200
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
DB 201 TCAATTTGACACCT---CCATCGATTGTAACAATATATCAACCATTTATGACGCA 257
QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAla 137
DB 258 GGTGCAAGAGTGCATTTGATTTCTGGGTTCTGTATCAGTAGAGGGGTATACAGGTAGT 317
QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsn---AspMetThrIle 156
DB 318 TATTCATCAGACACTGCTTCATATTCCTTTCATTCATCCTCAGATTTTCATCAT 377
QY 157 LeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSer-----AlaProGly 174
DB 378 CTTAGAGTGTCTGGTAAC-----AACGAGCAATCACTATCACTACTCACTCAATCC 428
QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg----- 189
DB 429 ACTGCAAGAATGTTATTACCTGGTCTCATCAACAATTCATGAAATTTTAACT 488
QY 190 -----ProSerPheGlySerTyrAlaAspAsnIleAsn-----HisValAla 203
DB 489 GATGTCCTCAATATATAAATATCAATCACTGTCGATATAAATCAAGATTTAATGT 548
QY 204 GlnPheSerSerArg 208
DB 549 GATTTCGATAGCAGA 563

RESULT 5
TA319G10P 532 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 319g10, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL492464
VERSION AL492464.1 GI:11867408
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 532)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing

```

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of *Trypanosoma brucei* (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

Source
1..532
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="319g10"

ORIGIN

Alignment Scores:
Pred. No.: 1.25e-11 Length: 532
Score: 216.50 Matches: 60
Percent Similarity: 48.37% Conservative: 29
Best Local Similarity: 32.61% Mismatches: 68
Query Match: 9.64% Indels: 27
DB: 29 Gaps: 6

US-09-985-689A-2 (1-434) x TA319G10P (1-532)

QY 93 ValPheGlnSerIleMetAspSerSerGlyGlyLeuGly----- 105
DB 2 GTAATGCCAAGTATAATGTGTGCGCAGGGGGGGAAGATTCTTCAGGGGTGGGGTGGC 61
QY 106 -----GlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPhe 119
DB 62 CATCCAGTCAGAGCTGTCTCTCCGCCACGAGTTACTCAATATTCTGTCGGGTATAT 121
QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaThr 139
DB 122 GCGCTGGAGCCGCTGTCTCTCAACTCGTGGGTTTGTGCTCCCTCCGAGTATCT 181
QY 140 ThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMetThrIleLeuPhe 158
DB 182 GCTGTGGAAGAAGATATGATGATGTTGCGAGTAGTATGACGATCGGCTATCTTC 241
QY 159 AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
DB 242 TCCACTGGCAACAGTATCCCAAGATGCG-----CTAATGACTCCGTGCTGTAAGAAC 295
QY 179 AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
DB 296 GTGATGTGGTGGGGTCACAAAAACGCG-----TTTGCAGCTTCGAAAGAC--- 343
QY 199 IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
DB 344 -----ATTGTTCTTCTCGTTTCTTCCCATGTTGTCACCAACATACGACGGTAGGAACCC 397
QY 219 AspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSer 238
DB 398 GATCTTGTGGTCCCGGGAAGAGGGTGTCTGCTCTTCTCTCTGCGAAAGCATCAGCT 457
QY 239 SerPheTrpAlaAsnHisAspSerTyrAlaTyrMetGlyGlyThrSerMetAlaThr 258
DB 458 AAA-----CAATGTAAGTGGTGGCCAGCGGNGTTTCATCGATGGCAACT 502
QY 259 ProIleValAla 262
DB 503 GCGGCGCTCGCG 514

```

RESULT 6
BQ142519/c
LOCUS
DEFINITION
Contig6 Metarhizium anisopliae sf. acridum ARSEF 324 Metarhizium
anisopliae var. acridum cDNA, mRNA sequence.
ACCESSION
BQ142519
VERSION
BQ142519.1
KEYWORDS
GI:20279578
SOURCE
Metarhizium anisopliae var. acridum
ORGANISM
Metarhizium anisopliae var. acridum
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic
Clavicipitaceae; Metarhizium.
REFERENCE
1 (bases 1 to 2141)
Freimoser, F.M., Screen, S., Baga, S., Hu, G. and St. Leger, R.J.
EST analysis of genes expressed by two different insect pathogenic
fungi during optimized secretion of proteins
Unpublished (2002)
JOURNAL
Contact: Freimoser F. M.
Department of Entomology
University of Maryland
4112 Plant Sciences Building, College Park, MD 20742, USA
Tel: 301 405 16 13
Fax: 301 314 92 90
Email: ff34@umail.umd.edu.
FEATURES
Location/Qualifiers
1..2141
/organism="Metarhizium anisopliae var. acridum"
/mol_type="mRNA"
/db_xref="ARSEF 324"
/db_xref="taxon:92637"
/cdone.lib="Metarhizium anisopliae sf. acridum ARSEF 324"
/notes="Vector: Unizap; Metarhizium anisopliae sf. acridum
was grown on insect cuticle and chitin for 24 hours. A
cDNA library was constructed in the unidirectional Lambda
vector Unizap."
ORIGIN
Alignment Scores:
Pred. No.: 9,18e-10 Length: 2141
Score: 208.50 Matches: 80
Percent Similarity: 41.72% Conservative: 41
Best Local Similarity: 27.59% Mismatches: 112
Query Match: 9.29% Indels: 57
DB: 13 Gaps: 13
US-09-985-689A-2 (1-434) x BQ142519 (1-2141)
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
1059 GGTGAGGGTACTTCGGTATATATATATGACACATGGTGTGGAGTCTCC----- 1012
Qy 41 SerMethisGluAlaPheArgGlyLeuValAlaLeuValAlaLeuGlyArgThrAsn 60
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
1011 -----CACCCCGAGTTTGGCGGTGCGCCACITGGCTCAGGAGCTTCATCAACGGTCAA 958
Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
957 AAC---CGTGATGCCCGCCAGCCATGGAGCTCCTCGCTGCTACTATT-----GGT 910
Qy 81 AlaThrAsnGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp--- 99
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
909 AGCCGAGCTACGGTGTGTCAAAATGCCAAGTCTTTCCTGCTCAAGTCTTCGTATGAC 850
Qy 100 ---SerSerGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla 118
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
849 CAGGCGCATGGTTCCTACTCGGTATCATCAGTGGCATGGACTTTGTTGCCAGGACTCC 790
Qy 119 PheSer-----AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
789 AAGAGTCGTACTGCCCCCAATGGCCACATGCTTCCATGATGCTGGGA----- 742

```

```

Qy 136 GlyAlaTyrThrAspSerArgAsnValAspTyrValArgGlyAsnAspMetThr 155
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
741 GGTGGTACTCGGGTCCGTCACACAGGGTGGCCCTGCTTTGGTCAGGTCTGGTGTCTTC 682
Qy 156 IleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
691 CTTCCGTCGGCGCTGGCAACAGATAACGGGATGCCCAAAACACCTCT---CCGCTTCC 625
Qy 176 AlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGly 193
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
624 GAGCTACTGCTGCACCTGTTGGTGGCCACTCGGTACAGATGACAGCCGATACCTTTTCC 565
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp 213
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
564 AACTAC-----
Qy 214 GlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSer 233
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
558 GGCAGAGT---GTGGATATCTTCGCTCTGGTACCGGCATTCTCTCCACC----- 511
Qy 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
510 -----TGGATCAAT-----GGCCGACCAACACCATCTCTGGC 478
Qy 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
477 ACCTCCATGGCTACTCCCATATTGCTGGT-----CITGCTGCCTACTTCAGT 430
Qy 274 LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
Db ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
429 GCTCTCAGCGCCAGAGCTAGCCCTGCGCTCTTTGCCAGAGATCCAGGACACTTCTACC 370
Qy 294 AlaAspValGlyLeuGlyTyrProAsnGly 303
Db ::::::::::: ::::::::::: ::::::::::: :::::::::::
369 AAGACGTGATCCCGCAATGTCCTGGCTGGC 340
RESULT 7
CA320325 771 bp mRNA linear EST 09-JUL-2003
LOCUS
DEFINITION
UI-M-FW0-cby-d-23-0-UI.r1 NIH BMAP_FW0 Mus musculus cDNA clone
IMAGE:6816072 5', mRNA sequence.
ACCESSION
CA320325
VERSION
CA320325.1
KEYWORDS
GI:24538449
SOURCE
Mus musculus
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 771)
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/
TITLES
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5
Location/Qualifiers
1..771
/organism="Mus musculus"
/mol_type="mRNA"
/strains="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6816072"
/tissue_type="whole brain"
/dev_stages="embryo 13.5,14.5,16.5,17.5dpc"

```

/lab_host="DH108 (T1 phage resistant)"
/clone_lib="NIH_BMAP_FW0"
/note="Organ: Brain; Vector: pX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is ACGGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN

Alignment Scores:
Pred. No.: 5,696-10 Length: 771
Score: 203.00 Matches: 84
Percent Similarity: 41.81% Conservative: 36
Best Local Similarity: 29.22% Mismatches: 109
Query Match: 9.04% Indels: 58
DB: 14 Gaps: 13

US-09-985-689A-2 (1-434) x CA320325 (1-771)

Qy	8	ValLysAlaAspValAlaGlnSerSerTyGlyLeuTyGlyGlnGlyValAla	27
Db	40	CTGCGGCGAGATGCTGTGG---CAGATGGATACACAGGTGTAATGTCAGAGTTGCT	96
Qy	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47
Db	97	GTTTGTGATCTACTGGGCTC-----AGTGAGACATCCGCAATTTTAAG	138
Qy	48	GlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn	63
Db	139	AAT-----GTGAGGAGAGAACCACTGACCAATGACGCGACCCCTG	180
Qy	64	---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThr	82
Db	181	GATGATGGCTAGGCCATCGCACATTCGTTGCAGT---GTGATTGCCAGCATGAGGGAG	237
Qy	83	AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSer	101
Db	238	TGCCAAGGATTTCCTCCAGATGCAGAGCTGCACATCTTCAGGGTCTTTTACCACCAATCAG	297
Qy	102	GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla	121
Db	298	-----GTGCTTACACATCTTCTGCTTCTGATGCTTCACTAT	336
Qy	122	GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThrAsp	141
Db	337	GCCATCTCTAAGAGATGACGCTTCTCAACCTTAGCATCGTGGGCCGCACTTCATGGAT	396
Qy	142	SerArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAla	159
Db	397	CATCGCTTGTGACAGGTGGGAATTAACAGCTACCAATGTAATATGTTCTGCT	456
Qy	160	AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla	179
Db	457	ATTGGCAATGATGGACCTCTCTATGGCACTCTGAATAACCTGCTGTGATCAGATGATGTG	516
Qy	180	IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIle	199
Db	517	ATTGAGTGGGGGCAATTGAC-----TTTGAAGAATAACATC	552
Qy	200	AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp	213
Db	553	-----GCTCGCTTTCTTCCAGGGGAATGACTACTCTGGGAATTACCAGGAGGCTAT	603

Qy	214	GlyArgIleLysProAspValMetAlaProGlyThrTyIleLeuSerAlaArgSerSer	233
Db	604	GGTCGTGTGAAGCTGACATTGTC-----ACCTATGGTCTCGAGTGGCGGTTCC	654
Qy	234	LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyAlaTyMetGlyGly	253
Db	655	GGTGTGAAGGGGGCTGC-----CGTGCACTCTCAGGG	687
Qy	254	ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal	273
Db	688	ACAGTGTGCTCTCCCGAGTGGTGGGGCGTCCCTTGTAGTANGCACAGTACAG	747
Qy	274	LysAsnArgGlyIleThrPro	280
Db	748	AAGCGGGAGCTGGTGAATCCT	768

RESULT 8
AQ652212/c
LOCUS
DEFINITION
Sheared DNA-8P2.TF Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-8P2, genomic survey sequence.
ACCESSION
AQ652212
VERSION
AQ652212.1 GI:5145398
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei
ORGANISM
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
1 (bases 1 to 508)
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
Doneelson,J., Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other GSSs: Sheared DNA-8P2.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nleayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tldb/mdb/tbdb/
Seq primer: M13-Forward
Class: shotgun.
FEATURES
Location/Qualifiers
1..508
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-8P2"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barell, Oxford University
Press, 1999)."

ORIGIN
Alignment Scores:
Pred. No.: 5,596-10 Length: 508
Score: 200.00 Matches: 57

```

Percent Similarity: 50.30%      Conservative: 28
Best Local Similarity: 33.73%    Mismatches: 56
Query Match: 8.91%              Indels: 28
DB: 28                          Gaps: 7

US-09-985-689A-2 (1-434) x A0652212 (1-508)

QY 156 IleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
DB 503 CTATCTTCTCCACTGGCACAGTATCCAGATGGC-----CTAATGACTCCGTTGCT 450
QY 176 AlaLyAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyr 195
DB 449 GGTAAAGAACGTGATGCGTGGGGTCCACAAAACGTG-----TTTACGCTTCG 399
QY 196 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg 215
DB 398 AAGAC-----ATTGTTCTTCTGCTTTCTTCGATGGTCCACATACGCGGTAGG 348
QY 216 IleLyProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerLeuAla 235
DB 347 ATGAACCCGATCTTGTGCGTCCGGGGAAGAGGTGTCTCTCTCTCTCTGCAAA 288
QY 236 ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSer 255
DB 287 GCATCAGCAAA-----CAATGTAAAGTGTGTGCCCAAGCGGGTTCATCG 243
QY 256 MetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 274
DB 242 ATGCAACTGGCGCGTGGCGGCGGCTACACTGTTGCGTCAGTATGTACGAGGCTA 183
QY 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuLeu----- 290
DB 182 AATCGAACTGCTTCACTTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123
QY 291 -----AlaGlyAlaAlaAspValGlyLeuGlyTyr----- 300
DB 122 CCGTTGAGTAACCTCCACGCTGAGTGGCTTTGGCGTCTTGATTATCTCTCTCTTTC 63
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArg 309
DB 62 CCAACGGGACACAGGGGATGTTCCGC 36

RESULT 9
LOCUS BJ387574 601 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ387574 Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION BJ387574.1 GI:19296958
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 601)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. 601
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44699"
/clone="dds3a18"
/sex="mat A"

FEATURES
source
1. 601
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44699"
/clone="dds3a18"
/sex="mat A"

/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores: 8.42e-10 Length: 601
Pred. No.: 199.50 Matches: 59
Score: 199.50 Conservative: 24
Percent Similarity: 44.86% Mismatches: 63
Best Local Similarity: 31.89% Indels: 39
Query Match: 8.89% Gaps: 6
DB: 12

US-09-985-689A-2 (1-434) x BJ387574 (1-601)

QY 198 AsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgGlyLys 217
DB 51 AATGAGAATAATATTTGTTTCATTCATCAAAAGGTCCACACATGATGTAGATGAA 110
QY 218 ProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerLeuAla----- 235
DB 111 CCTGATTAGTTGCCCCCTGGTGAATATATTACATCGGCAAGATCAATGTTGCAATACA 170
QY 236 -----ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMet 251
DB 171 ACAGACCAATGTGGTGTGCTCTTTA---CCAATACAAATGCATTATTGGCG---ATA 224
QY 252 GlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHis 271
DB 225 TCTGTATCATCAATGGCAACCTCATTTGCGAGCAGCAGCAACAACAATCTTAGACAATAT 284
QY 272 -----PheValLysAsnArgGlyIleThrProLys 281
DB 285 TTAGTTGATGTTATTATCCAACTGGTTCAATTCTAGATCAACAATAAATAATACACCAACT 344
QY 282 ProSerLeuLysAlaAlaLeuLeuAlaGlyAla----- 293
DB 345 GGATCAATTATAAAGCATTAATGATTAAATGCTCAGTTATTAAATGGTACATTTCAA 404
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsn----- 304
DB 405 TTGATTACATCATCAAGTATTACATATCCATCAACCAAGTTTTTGAAATTTTGCAGGT 464
QY 305 -----GlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 465 GCAAGTTAGTTCAGGTGGGTGCTATTAGATAGTAAATGTTGTTACATGTTGTCAT 524
QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
DB 525 AATAATAATAGTAATAATAATAAACAACATCAGTATGATACTAAATTTGTTGGTATT 584
QY 341 AlaGlyLysProLeu 345
DB 585 GGTGGATTAGATTTA 599

RESULT 10
LOCUS BJ369190 633 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ369190 Dictyostelium discoideum cDNA library, CF Dictyostelium
ACCESSION BJ369190.1 GI:19278573
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 633)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics

```

1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES

source
1..633
/organism="Dictyostelium discoideum"
/mol_type="rRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc49116"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Alignment Scores:
Pred. No.: 2,66e-09 Length: 633
Score: 195.00 Matches: 58
Percent Similarity: 41.5% Conservative: 28
Best Local Similarity: 28.0% Mismatches: 49
Query Match: 8.6% Indels: 72
DB: 12 Gaps: 6

US-09-985-689A-2 (1-434) x BU369190 (1-633)

```

Qy 136 GlyAlaTyrThrAspSerArgAsnValaAspTyrValaTgLyS---AsnAspMet 154
Db 6 GGTGGTTATTCGATGATGCTGGTGTATGATGATTCCTCTATGATACCCAGAAATC 65
Qy 155 ThrLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGly 174
Db 66 TCTATACTAAGAGCTGCTGGTAAT--AACGAGCTATTTCGATCTTTATTAGTCAAGCA 122
Qy 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
Db 123 ACAGCTAAATGCAATACATGCTGGTGTGCTGAGCAACAGCTCATGTAAATATTGTGPCA 182
Qy 195 -----TyrAlaAspAsnIle----- 199
Db 183 GATGCATTGGAATATTGATGATTCCTCAGATATGCTAATTTCAAGACCATGTTATTC 242
Qy 199 ----- 199
Db 243 GATAAGAAGTATGTAAATTATACGACCGCTAAATGTTGCTCAGAGGTTTCAATGTTAAA 302
Qy 199 ----- 199
Db 303 GGTTCACAAATTATGTTGCCGATCTATTAAACAAATGCAATCGATTCGATTCACAACA 362
Qy 200 -----AsnHisValAlaGlnPheSerSerArgGlyProThrLys 212
Db 363 CAACTCAATTTTATAATGAAATAATATGCGATCATTTCTCATCAAGGGTCCACACAT 422
Qy 213 AspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer 232
Db 423 GATGGTAGATTGAACCTGATATAGTTGCACTGTTGATATATTATACATCGGCAAGATCA 482
Qy 233 Ser-----LeuAlaProAspSerSerPheThrAlaAsnHisAspSer 246
Db 483 AATGGTGAGAAATTCACAGACCAATGTTGGTGTATGTTCTTTA-----CCAAATGCAAT 536
Qy 247 LysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnVala 266
Db 537 GGTTCAATGTCATCTCTGGTACATCATGCGCAACACCATTTGGCAACAGCAGCAACA 596
Qy 267 GlnLeuArgGluHisPheVal 273
Db 597 ATTCCTTAGCAATATTAGTT 617

```

RESULT 11

BI750157
LOCUS

BI750157 718 bp mRNA linear EST 25-SEP-2001

DEFINITION

Fg02_10g08_R Fg02_AAFc_ECORC_Fusarium graminearum mycelium
Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence.

ACCESSION

BI750157

VERSION

BI750157.1 GI:15771959

KEYWORDS

EST

SOURCE

Gibberella zeae

ORGANISM

Gibberella zeae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

REFERENCE

1 (bases 1 to 718)

AUTHORS

Harris, L.J., Glassco, T., Rocheleau, H., Allard, S., Chapados, J.,
Couroux, P., De Moors, A., Hattori, J.I., Ouellet, T., Robert, L.S.,
Singh, J.A., Sprott, D. and Tinker, N.A.

TITLE

Expressed Sequence Tags from Fusarium graminearum mycelium

JOURNAL

Unpublished (2001)

COMMENT

Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-Food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6586
Email: harrisj@em.agr.ca.

FEATURES

Location/Qualifiers

1..718

/organism="Gibberella zeae"

/mol_type="rRNA"

/strain="DAOM 180378"

/db_xref="taxon:5518"

/clone="Fg02_10g08"

/tissue_type="Mycelial tissue"

/dev_stage="Assexual"

/lab_host="E. coli (Sure cells)"

/note="Vector: Bluescript SK+/XhoI-ECOR1; Site_1: EcoRI;
Site_2: XhoI; Mycelial tissue was collected from V8 agar
plates after a growth period of 6-7 days at 25 C with 14
hrs (FL/UV) day light exposure. Mycelia was ground in
liquid nitrogen prior it's storage at -80 C until RNA
extraction. Directional cloning with 5' end of cDNA cloned
into EcoRI site of pBluescript and 3' end of cDNA cloned
into XhoI site of pBluescript (Stratagene, La Jolla, CA)."

ORIGIN

Alignment Scores:

Pred. No.: 3,3e-09 Length: 718
Score: 195.00 Matches: 67
Percent Similarity: 43.6% Conservative: 30
Best Local Similarity: 30.1% Mismatches: 89
Query Match: 8.6% Indels: 36
DB: 12 Gaps: 10

US-09-985-689A-2 (1-434) x BI750157 (1-718)

```

Qy 46 PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThr 65
Db 9 TTCAGGGTCTGCTCAGGCTGTCTACACTGCTTCAGCGCCAGACCGCT---GACACC 65
Qy 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-----GlyAlaThrAsn 83
Db 66 AACGGTCACGGAACCTCAGCTTGTGCACTATTTCGCGAAAGACATACGTTGTGCCAAG 125
Qy 84 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGly 103
Db 126 AAGGCCACCATC---CAAGCTGTCAAGGTTCTCCAGGTT-----AGTTCAATCCAGC 173
Qy 104 LeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAla 123
Db 174 ACCTCCATCATCTCGTGGCTTCAACTGGGCTGCAACGACATCATCTCCAAGGGC--- 230
Qy 124 ArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrAspSerArg 143
Db 231 CGAACCAAGACACTCAGTCGTCATATGTCTCTCGCGGGTGGTTACTCTGCTCTCTCAAC 290

```

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source
1. 1002
/organism="Zygosaccharomyces rouxii"
/mol_type="genomic DNA"
/strain="CBS 732"
/db_xref="taxon:4956"
/clone_lib="AROAA010H04"
/notes="end : T3"
misc_feature
2. .>979
/notes="similar to Saccharomyces cerevisiae ORF YCR045c [similarity to serin proteases]"
/evidence="not_experimental"

ORIGIN

Alignment Scores:
Pred. No.: 8.38e-09 Length: 1002
Score: 193.50 Matches: 80
Percent Similarity: 41.90% Conservative: 39
Best Local Similarity: 28.17% Mismatches: 102
Query Match: 8.62% Indels: 63
DB: 29 Gaps: 14

US-09-985-689A-2 (1-434) x CNS06D6B (1-1002)

QY 14 GlnSerSerTyrGlyLeuTyr-----GlyGlnGlyGlnIleValala 27
Db 278 GAAGAGAGTTTCATTATTTACTATTATAATGGCACAGGGTAGAATAATCAATGCATAT 337
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 338 ATWATTGATCTCTGGAATC-----TATAAGAACACATAAGGACTTCTGT 379
QY 48 GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAla-----AsnAspThrAsn 66
Db 380 GGAGCT-----GCATATTGGCGGAGAGCTTTACAGGGGAGAGTCTCGTGATCGTAAT 433
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsnLysGlyMet 86
Db 434 GGACATGGAACCTACGTTGCTGGCATTTGTG-----GGATCGAGCAACTTTGGTGT 484
QY 87 AlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyGly 106
Db 485 GCTAAAGAGTGAATTTAATTGACGTTAAAGCTTTGACCAATAGAGC-----GlnAlaPhe 119
QY 107 LeuProSerAsnLeuGlnThrLeuPheSer-----GlnAlaPhe 119
Db 533 ---CAAGGCAATTTAACACACAGTAATTAGCCTCTTGAATTTGCGTGAATTCATTTGTA 589
QY 120 SerAlaGlyAlaArg-----IleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAla 137
Db 590 TCATCGGGCAAAAGGGTGGTGGCCATCTTTTCATTTGGTGGCCGACAGAACAGTGA 649
QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeu 157
Db 650 ATAATATCAAGCAATTTAAA-----GCTGCTCATGAAGCAGGCGCTCATTTATGTA 697
QY 158 PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
Db 698 GTTGAGGAGGTAAATCTTAATTAAT---GCATGTGGGAATAGCCCTGCCCTGCACCT 754
QY 178 AsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAsp 197
Db 755 GAGGCTATTACTTAGGGGCC-----TTTGATGAT 784
QY 198 AsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLys 217
Db 785 CGTACAGATACAATTGCAAAATTTAGTAATTTGGGGTCCATGT-----826

QY 144 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 163
Db 291 AACGCTGTGAGTCTKCYTCCAGCTCCGGTATTATCTGTGCAATTCGTCGCGGTAAACGAT 350
QY 164 GlyProAsnGlyGlyThrIleSerAlaPheGlyThrAlaLysAsnAlaIleThrValGly 183
Db 351 GGTGCCAACGCTGCCAACACTTCT---CTGCTCTCTCTCCAGCGCCCAACACTGTCGT 407
QY 184 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValala 203
Db 408 GCCATTGACAGC-----AACTGGGC-----ATTGCC 434
QY 204 GlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaPro 223
Db 435 TCGTACTCAACTACGCTACCGTCTCTC-----GATATCTTTGCYCT 476
QY 224 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsn 243
Db 477 KGCACACAGGTTCTCTCCGCC-----TGGTACACC 506
QY 244 HisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 263
Db 507 AGCAACAGTGCACCAACACCATCAGCGKALSTCCTGCTACTCTCCACATTCGCCGA 566
QY 264 AsnVal 265
Db 567 CTGTGC 572

RESULT 12

CNS06D6B 1002 bp DNA linear GSS 14-JUN-2001
LOCUS T3 end of clone AROAA010H04 of library AROAA from strain CBS 732 of
Zygosaccharomyces rouxii, genomic survey sequence.

ACCESSION

AL393417

VERSION

AL393417.1 GI:12143597

KEYWORDS

GSS.

SOURCE

Zygosaccharomyces rouxii

Zygosaccharomyces rouxii

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.

1 (bases 1 to 1002)

Souciet J.L., Aigle, M., Artiguenave, F., Blandin, G.,

Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,

de Montigny, J., Dujon, B., Durrien, P., Lepingle, A., Llorente, B.,

Maupertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Tekalia, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,

Wincker, P. and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FBES Lett. 487 (1), 3-12 (2000)

11152876

2 (bases 1 to 1002)

de Montigny, J., Straub, M., Potier, S., Tekalia, F., Dujon, B.,

Wincker, P., Artiguenave, F. and Souciet, J.

Genomic exploration of the hemiascomycetous yeasts: 8.

Zygosaccharomyces rouxii

FBES Lett. 487 (1), 52-55 (2000)

20584718

11152883

3 (bases 1 to 1002)

Genoscope.

Direct Submission

Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,

2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

secre@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

COMMENT

QY 218 ProaspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerSerLeuAlaProasp 237
 Db 827 GTGATATTTCATCATGCGCTAAAGTAAATCATTCG---TCGCATTCCCACT--- 880
 QY 238 SerSerPheTrpAlaAsnHisaspSerlystYrAlaTyMetGlyGlyThrSerMetAla 257
 Db 881 -----CATAAACCAATTGCATTTCAGGTACTTCGATGCT 916
 QY 258 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVallyAsnArgGly 277
 Db 917 TCACCAAGGTTACTGGACTTCTGCTATATTATTGGATAAAGGCGTTCAACCAAGAAAT 976
 QY 278 IleThrProLys 281
 Db 977 ATTAAAGCTAAG 988

RESULT 13
 BQ770462
 LOCUS
 DEFINITION
 UI-M-F10-bvV-m-19-0-UI.r1 NIH BMAP_F10 Mus musculus cDNA clone
 IMAGE:5702970 5', mRNA sequence.
 BQ770462.1 GI:21978936
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 665)
 NIH-MGC http://mgs.nsl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. .665
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5702970"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP F10"

/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Benaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is CAGCCACGAC. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

ORIGIN

Alignment Scores:

Pred. No.: 1,5e-08 Length: 665
 Score: 188.00 Matches: 71
 Percent Similarity: 44.58% Conservative: 36
 Best Local Similarity: 29.58% Mismatches: 91
 Query Match: 8.37% Indels: 42
 DB: 13 Gaps: 10
 US-09-985-689A-2 (1-434) x BQ770462 (1-665)
 QY 63 AsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThr 82
 Db 14 GATGATGGGTAGCCATGCGCATTGCTTGCAGGT---GTGATTCGCCAGCATGAGGAG 70
 QY 83 AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSer 101
 Db 71 TGCCAGGATTGCTCCAGATGCAGAGCTGCACATCTTCAGGGCTTTTACCAACATCAG 130
 QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
 Db 131 -----GTGCTCTACACATCTTGGTTTCTGGATGCTTCAACTAT 169
 QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThrAsp 141
 Db 170 GCATCTCTAAAGAGATGGACGTTCTCACTTAGCATCGTGGGCCGCGCTTCATGGAT 229
 QY 142 SerArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAla 159
 Db 230 CATCCGTTTCTCACAAGGTGTGGGAATTAACAGCTAACAAATGTAATATTATGTTCTGCT 289
 QY 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
 Db 290 ATTGGCAATGATGACCTCTCTATGGCATTCTGTAACCTCTGATCATGATGATGATG 349
 QY 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIle 199
 Db 350 ATTGGAGTGGTGGCATTCGAC-----TTTGAAGATAAACATC 385
 QY 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp 213
 Db 386 -----GCTCGCTTTCTCCAGGGGAATGACTACTCTGGGAATTACAGGAGGCTAT 436
 QY 214 GlyArgIleLysProAspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerSer 233
 Db 437 GGTCTGTGAAGCTGCATTGTC-----ACCTATGGTCTGGAGTGGGGGTCTCC 487
 QY 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyMetGlyGly 253
 Db 488 GGTGTGAAGGGGGCTGC-----CGTGCACCTCTCAGGG 520
 QY 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
 Db 521 ACCAGTCTCGCTTCNCAGTGGTGGCGCGCTCACCTTGTGTAGTACGACAGTACAG 580
 QY 274 LysAsnArgGlyIleThrProLysProSerLeuLysAlaAlaLeuIleAlaGlyAla 293
 Db 581 AAGCGGAGCTGGTGT-----AATCTGCCAGTGTGAAGCAGCTTTGTAGCTCAGCC 634
 RESULT 14
 BZ893395/c
 LOCUS
 DEFINITION
 H12_0177 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
 genomic survey sequence.
 BZ893395
 VERSION BZ893395.1 GI:33343985
 KEYWORDS
 SOURCE
 ORGANISM
 Halorubrum lacusprofundi
 Halorubrum lacusprofundi
 Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 Halobacteriaceae; Halorubrum.
 1 (bases 1 to 716)
 Gao, Y., Roach, J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M.,
 DasSarma, S., Ng, W.V. and Hood, L.
 Low-pass Sequencing for Microbial Comparative Genomics
 TITLE

JOURNAL COMMENT

Unpublished (2003)
Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel.: 206 732 1412
Fax: 206 732 1299
Email: ygoo@systemsbiology.org
Seq primer: M3 Forward
Class: shotgun.

FEATURES
source
Location/Qualifiers
1..716
/organism="Haloarubrum lacusprofundi"
/mol_type="genomic DNA"
/strain="ATCC 49239"
/db_xref="taxon:2247"
/clone_lib="H1 pUC18 Library"
/note="Vector: pUC18; Site 1: SmaI; A shotgun library was constructed from Haloarubrum lacusprofundi genomic DNA using pUC18/SmaI/BAP plasmid"

ORIGIN
Alignment Scores:
Pred. No.: 2,73e-08 Length: 716
Score: 186.00 Matches: 70
Percent Similarity: 47.73% Conservative: 35
Best Local Similarity: 31.82% Mismatches: 78
Query Match: 8.29% Indels: 38
DB: 28 Gaps: 10

US-09-985-689A-2 (1-434) x BZ893395 (1-716)

QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
.....
Db 701 TCGAGTCGTGGACGNWAGCGGCACGAACTCAGTCCGCCGGAACGTCGCGGCGCGA 642

QY 80 GlyAla-----ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
.....
Db 641 AACCGAGTCGGACCGCAATCGCGTCGACCACGACGCTCTCCACGGGTATCAAGTG 582

QY 98 MetAsp-SerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
.....
Db 581 TTCGACGACACGGGA-----CCAAGCACGCTTCGTCGCGTCGTCGCG 537

QY 117 nalaPheSerAlaGlyAlaArgIleHisThrAsnSerTrp-----GlyAlaAlaVa 134
.....
Db 536 GGAATGGAACACGCGACGAGGATCGGACGTCGACGCTTCAGATGAGCCCTCGTCG 477

QY 134 lAsnGlyAlaTyThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMe 154
.....
Db 476 GACGGGACTTACCATCTTCATCGAACCG-GTTCCGAC---ACTCGAGTCCCGGNA 421

QY 154 tThrLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGln 174
.....
Db 420 GATCGCGCTGTTTCGGCGCGGACATCGGTCAA-----GGAACGTCGAGTTCCTCCG 367

QY 174 yThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg-----ProSerPh 192
.....
Db 366 GAACGCTACGACTCGCTCGGTCGGAGCGGTCAACGACGCGCGGCGTCCCGACTT 307

QY 192 eGlySerTyAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThr-- 211
.....
Db 306 CTCAGC---GGAGAGACGATCAACACGTCACGCGCGTGGGAGAGGATGCCCCCCCGGA 250

QY 212 ----LysAspGlyArgIleLysProAspValMetAlaProGlyThrTyIleLeuSerAl 230
.....
Db 249 CTGCCCCGACGAGTACGTGTTCCGACGCTGTCGGCCGACGCGTCCGAGGTATATTCTCG 190

QY 230 aArg-----SerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLy 247
.....
Db 189 GGAAACCGCGCGGTACAACATCCGAAGGAC-----

QY 247 sTyAlaTyMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 266
.....

Db 158 -----GGCACCTCCATGGCCGACGACGACGTCAGCGCGTTCGCGGCG 117

RESULT 15
CF727824
LOCUS
DEFINITION
UI-M-HB0-cl-m-24-0-UI.r1 NIH_BMAP_HB0 Mus musculus cDNA clone
IMAGE:30550487 5', mRNA sequence.
CF727824
CF727824 GI:37601992
EST.
Mus musculus (house mouse)
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CF727824 675 bp mRNA linear EST 09-OCT-2003
UI-M-HB0-cl-m-24-0-UI.r1 NIH_BMAP_HB0 Mus musculus cDNA clone
IMAGE:30550487 5', mRNA sequence.
CF727824
CF727824 GI:37601992
EST.
Mus musculus (house mouse)
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 675)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pyX-5,
Location/Qualifiers
1..675
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30550487"
/tissue="eye"
/dev_stage="embryo 12.5,13.5,14.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HB0"
/notes="Organ: Eye; Vector: pyX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996 Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor , digested with NotI and then cloned
directionally into pyX-Asc vector . The library tag
is "TATTCAAGT". This library was created for the University
of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

ALIGNMENT SCORES
Pred. No.: 3.12e-08 Length: 675
Score: 185.00 Matches: 70
Percent Similarity: 44.54% Conservative: 36
Best Local Similarity: 29.41% Mismatches: 30
Query Match: 5.24% Indels: 42
DB: 14 Gaps: 10

US-09-985-689A-2 (1-434) x CF727824 (1-675)

QY 63 AspAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThr 82
.....
Db 25 GATCATGGCTAGCCATGGCATTCGTTTCAGGT---GTGATTCGCCAGCATGAGGAG 81

QY 83 AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSer 101
.....
Db 82 TGCAAGGATTTCTCCAGTGCAGATCGACATCTTCAGGTCTTTTACCACATCATG 141

```
QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
Db      :|||
142 -----GTGCTTACACATCTTGGTTTCTGGATGCTTCAACTAT 180
QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThrAsp 141
Db      :|||
181 GCCATCCTAAGAAGATGGACGTTCTCAACCTAGCATCGGTGGGCCGACCTCATGGAT 240
QY 142 SerArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAla 159
Db      :|||
241 CATCGTTTGTTCACAAAGGTGGGAATTAACAGCTAACCAATGAATATGTTTCTGCT 300
QY 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
Db      :|||
301 ATTGCAATGATGGACCTCTCTATGGCACTCTGAATACCCCTGCTGATCAGATGGATGTG 360
QY 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIle 199
Db      :|||
361 ATTGGAGTGGTGGCATTGAC-----TTTGAAGATAACATC 396
QY 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp 213
Db      :|||
397 -----GCTCGCTTTTCTCCAGGGAATGACTACCTGGGAATTACAGGAGCTAT 447
QY 214 GlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSer 233
Db      :|||
448 GGTGCTGTGAAGCCTCACATTGTC-----ACCTATGGTGTGGAGTGGGGGTCC 498
QY 234 LeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253
Db      :|||
499 GGTGTGAAGGGGGCTGC-----CGTGCACCTCTCAGGG 531
QY 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
Db      :|||
532 ACCAGTGTGCTTCCCGAGTGTGCTGGGGCGTACCTTGTAGTAAGCACACGTACAG 591
QY 274 LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAla 291
Db      :|||
592 AAGCGGAGTGGTG-----AATCCTGCCAGTGTGAAGCAAGCTTTGATAGCG 639
```

Search completed: April 4, 2004, 11:54:37
Job time : 2290.84 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:47:19 ; Search time 3544.27 Seconds
(without alignments)
5295.174 Million cell updates/sec

Title: US-09-985-689A-3
Perfect score: 2263
Sequence: 1 NDVARGIVKADVAQNNGLY.....EVQAINVPSGQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-O=/cgn2_1/USPTO_spo3/US09985689/runat_31032004_161806_4155/app_query_fasta_1.3498
-DB=GenEmbl -QEXT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09985689@cgn_1_11925@runat_31032004_161806_4155 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORE=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPT=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXI=7

Database : GenEmbl :
1: gb_ba :
2: gb_htg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_to :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vi :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_on :
21: em_or :
22: em_ov :
23: em_pat :
24: em_ph :
25: em_pi :
26: em_ro :
27: em_sts :
28: em_un :

29: em_vi :
30: em_htg_hum :
31: em_htg_inv :
32: em_htg_other :
33: em_htg_mus :
34: em_htg_pin :
35: em_htg_rod :
36: em_htg_mam :
37: em_htg_vrt :
38: em_sy :
39: em_htgo_hum :
40: em_htgo_mus :
41: em_htgo_other :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2263	100.0	1299	1	AB046402	AB046402 Bacillus
2	2251	99.5	1299	1	AB046405	AB046405 Bacillus
3	2239	98.9	2218	6	E03808	E03808 DNA encodin
4	2234	98.7	1299	1	AB046404	AB046404 Bacillus
5	2040	90.1	3003	6	AR069954	AR069954 Sequence
6	2040	90.1	3003	6	BD062155	BD062155 Nucleic a
7	2020.5	89.3	1302	1	AB046406	AB046406 Bacillus
8	1999.5	88.4	1923	1	AB084155	AB084155 Bacillus
9	1998.5	88.3	1920	1	AB046403	AB046403 Bacillus
10	1998.5	88.3	1920	6	AR368116	AR368116 Sequence
11	1995.5	88.2	1923	6	AR368118	AR368118 Sequence
12	1994.5	88.1	1305	6	AX829476	AX829476 Sequence
13	1994.5	88.1	1923	6	AB051423	AB051423 Bacillus
14	1994.5	88.1	1923	6	AR368117	AR368117 Sequence
C 15	612.5	27.1	60006	1	AF268611	AF268611 Unculture
C 16	559.5	24.7	301205	1	AE017218	AE017218 Geobacter
17	502	22.2	5890	3	DDU60086	DDU60086 Dictyosteli
18	468	20.7	5772	3	AF466309	AF466309 Dictyosteli
19	463.5	20.5	6115	3	DDU20432	DDU20432 Dictyosteli
20	443	19.6	1977	6	AR201152	AR201152 Sequence
21	442.5	19.6	5109	3	AF263455	AF263455 Dictyosteli
22	422.5	18.7	1236	6	AR201146	AR201146 Sequence
23	422.5	18.7	1962	6	AR201155	AR201155 Sequence
C 24	422.5	18.7	12452	1	AE010365	AE010365 Pyrococcus
C 25	405	17.9	126928	2	AC096673	AC096673 Trypanoso
C 26	399.5	17.7	14376	1	AE013049	AE013049 Thermoana
C 27	396.5	17.5	303450	1	SC093913	SC093913 Streptomy
C 28	377.5	16.7	299850	1	AP004601	AP004601 Oceanobac
29	348.5	15.4	2539	6	AR202322	AR202322 Sequence
30	347.5	15.4	1239	1	AF305633	AF305633 Thermoana
31	347.5	15.4	1239	1	AY028704	AY028704 Thermoana
32	346.5	15.3	135638	1	AF484556	AF484556 Streptomy
33	345.5	15.3	3348	1	D83672	D83672 Streptomyces
C 34	345.5	15.3	12295	1	AB013026	AB013026 Thermoana
35	344	15.2	292550	1	AP001513	AP001513 Bacillus
36	343.5	15.2	2809	6	AR202321	AR202321 Sequence
37	343.5	15.2	3900	1	AB007809	AB007809 Streptomy
38	336	14.8	1329	6	AX433519	AX433519 Sequence
C 39	336	14.8	300800	1	SC093912	SC093912 Streptomy
40	334	14.8	2731	1	AF015325	AF015325 Thermococ
41	329	14.5	6854	1	BSU39230	BSU39230 Bacillus sp
42	329	14.5	300425	1	AP005044	AP005044 Streptomy
43	327	14.4	291000	1	SC093910	SC093910 Streptomy
C 44	327	14.4	302300	1	AP005034	AP005034 Streptomy
C 45	321	14.2	301150	1	AP004602	AP004602 Oceanobac

ALIGNMENTS

RESULT 1

Db	1261	CCATCTGACACACAGCGTTTCTCACTAGCTATCGTACAT	1299	
RESULT 2				
AB046405		1299 bp	DNA	linear
LOCUS				BCT 23-JAN-2001
DEFINITION				Bacillus sp. SD521 PROD gene for protease, partial cds.
ACCESSION				AB046405
VERSION				AB046405.1
KEYWORDS				GI:12381942
SOURCE				Bacillus sp. SD521
ORGANISM				Bacillus sp. SD521
REFERENCE				Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
AUTHORS				1 (sites)
TITLE				Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
JOURNAL				Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE				20568675
PUBMED				11118284
REFERENCE				2 (bases 1 to 1299)
AUTHORS				Saeki, K.
TITLE				Direct Submission
JOURNAL				Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail: 387185@kasanet.kao.co.jp, Tel: 81-285-68-7400, Fax: 81-285-68-7403)
FEATURES				Location/Qualifiers
source				1..1299
gene				/organism="Bacillus sp. SD521"
CDS				/mol_type="genomic DNA"
				/strain="SD521"
				/db_xref="taxon:133780"
				1..1299
				/gene="PROD"
				<1..>1299
				/gene="PROD"
				/codon_start=1
				/transl_table=11
				/product="protease"
				/protein_id="BAB21268.1"
				/db_xref="GI:12381943"
				/translation="NDVARGIVKADVAONNYGLYGGQGVAVADTGLDGRNDSMHE
				APRGIKITALYALGRNTNNDNGHGVAGVLGNALNKGWAPQANLVFOSIMDSGG
				LGGLPSNLNTLFSQAWNAGARIHNSWGPVNGAYTANSQVDEYVNDVTLFPAAG
				NEGNSGILSPGTAKXALITVQATENRPSFGSLADPNHIAQSSRGATGRKED
				VIAPGTFILSRSLAPDSFWANINSKIAFMGGTSMATPIVAGNVAAQLREHFINKRG
				ITPESLIKAALIGATDVGLGYPGGQGWGRVTLDKSLNVAYVNEATALATQKATY
				SFOAQAGPLKISLVWTDAPGSTTASYTLVNDLDELITAPNGQKYVGNDFSPYDNNW
				DGRNVENVFVNAPOSGTYTIEVQAYNVPSGPQRFSLAIVH"
ORIGIN				
Alignment Scores:				
Pred. No.:	1,77e-133	Length:	1299	
Score:	2251.00	Matches:	430	
Percent Similarity:	99.54%	Conservative:	1	
Best Local Similarity:	99.31%	Mismatches:	2	
Query Match:	99.47%	Indels:	0	
DB:	1	Gaps:	0	
US-09-985-689A-3 (1-433) x AB046405 (1-1299)				
Qy	1	AsnAspValAlaArgGlyLeuValValysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr	20	
Db	1	AATGATGTAGCAAGAGGAATAGTAAAGCAGCGTTGCACAAACAAATACGGACTATAT	60	
Qy	21	GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40	
Db	61	GGACAGGTCAGTAGTGTGACATCGCGATACGGGTTTAGATACAGGTCGTAACGATAGT	120	
Qy	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60	
Db	1201	GTATTTTAAACGCTCCGCAATCTGGAACGTATACATTTGAGGTTCAAGCGTATATGTA	1260	
Db	121	TCTATGCATGAAGCATTCGTCGGAAAATTACAGCTCTTTACGCGTTAGGAAGAACTAAC	180	
Qy	61	AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla	80	
Db	181	AATGCAAAATGATCCGAATGGCATGGTACGCATGTAGCAGGTCTGTGTACTTGGTAATGT	240	
Qy	81	LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer	100	
Db	241	TTAAATAAAGGAATGGCTCCGCAAGCTAATCTAGTCTTCCAACTCTATTATGATAGCAG	300	
Qy	101	GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla	120	
Db	301	GGAGGATTAGGTGGATTACCATCGAATTTGAATTCGTTATTATTTAGCAAGCTTGGAA	360	
Qy	121	GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyValaTyrThrAlaAsn	140	
Db	361	GGGCTAGAAATTCATACTAATCTTGGGTGCTCCAGTAAATGGAGCGTACACTGCTAAC	420	
Qy	141	SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly	160	
Db	421	TCGAGACAAGTGGATGAGTATGTTGAAATAATATGATATGACGCTACTTTTTCGACGAG	480	
Qy	161	AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr	180	
Db	481	AATGAAGTCTTAATTCAGAACAAATAGTGTCCAGGCACAGCAAAATATGCCATTAG	540	
Qy	181	ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis	200	
Db	541	GTCCGGCAACGGAATAATATGCCGAGCTTCGGTTTCATTAGCAGATAACCCAAATCAT	600	
Qy	201	IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr	220	
Db	601	ATCGCAAAATTTTCATCAAGAGAGGCTACGAGATGGAGCAATTAACACAGACGTAA	660	
Qy	221	AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp	240	
Db	661	GCTCCTCGAACAATTTATTTTATCAGCACGTTCTCTCTAGCCCCAGACTCTTCGTTT	720	
Qy	241	AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal	260	
Db	721	GCGAATTAACAGTAGTAAGTACGATCGTACATGGCGGTACCTCTATCGGCACACTAT	780	
Qy	261	AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys	280	
Db	781	CGCGGGAATGTCCGCAATTCACGAGCATTTTATAAAAAATAGAGGAATTACACCTAA	840	
Qy	281	ProSerLeuIleLysAlaAlaIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro	300	
Db	841	CCTTCCTTATAAAGCTGCACCTTATCGTGGGCTACTGATGTTGGTTTAGGATATCCA	900	
Qy	301	SerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal	320	
Db	901	AGTGGTGACCAAGGCTGGGGGGGTGTACTCTAGATAAATCCGTTAAATGACGGTATG	960	
Qy	321	AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla	340	
Db	961	AATGAAGCAACTGCATTAGCAACAGACAAAAAGCAACGTTATTCGTTCCAGCACAA	1020	
Qy	341	GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer	360	
Db	1021	GGTAAACCTTTTAAATCTCGTTAGTATGGACAGATGCACCTCGGAAGTAGTACACT	1080	
Qy	361	TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal	380	
Db	1081	TATACACTAGTTAATGATTAGTATCTAGTTATTACTGTCTCCGATGGACAAAAATAT	1140	
Qy	381	GlyAsnAspPheSerTyrProTyrAsnAsnTyrAspGlyArgAsnAsnValGluAsn	400	
Db	1141	GGAAATGATTATTAGTTATCTTATGATAATACTGGATGGTGGCAACAATGTTGAGAAC	1200	
Qy	401	ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal	420	
Db	1201	GTATTTTAAACGCTCCGCAATCTGGAACGTATACATTTGAGGTTCAAGCGTATATGTA	1260	

```

QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 CCACTCTGGCCACAGCGTTCTCCTACCTAGCTACGTACAT 1299

RESULT 3
EO3808
LOCUS          2218 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION    DNA encoding alkaline protease Ya.
ACCESSION     E03808
VERSION       E03808.1  GI:2172022
KEYWORDS      JP 1992197182-A/1.
SOURCE        Bacillus sp.
ORGANISM      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE     Tobe, S., Odera, M. and Asai, Y.
AUTHORS       DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE
TITLE         PROTEASE YA USING THE DNA
JOURNAL       Patent: JP 1992197182-A 1 16-JUL-1992;
              LION CORP
COMMENT       OS Bacillus sp.
              PN JP 1992197182-A/1
              PD 16-JUL-1992
              PF 28-NOV-1980 JP 1990327110
              PI TOBE SEIICHI, ODERA MOTOYASU, ASAI YOSHIO
              PC C12N15/57.C11D3/386; C12N9/54, (C12N15/57, C12R1.07), (C12N9/54,
              PC C12R1.07);
              CC strandedness: Double;
              CC topology: Linear;
              CC hypothetical: No;
              CC anti-sense: No;
              CC *source: strain-Y strain;
              CC *key: clone-pUB8A;
              FH Location/Qualifiers
              PH
              FT CDS      218..2125
              FT          /product='precursor of alkaline protease Ya'
              FT sig_peptide 218..823
              FT mat_peptide 824..2122
              FT          /product='Precursor of alkaline protease Ya'.

FEATURES
    source          Location/Qualifiers
    1..2218
    /organism="Bacillus sp."
    /mol_type="genomic DNA"
    /db_xref="taxon:1409"

ORIGIN
Alignment Scores:
Pred. No.:      1.9e-132      Length:      2218
Score:          2239.00      Matches:     428
Percent Similarity: 99.31%      Conservative: 2
Best Local Similarity: 98.85%      Mismatches:  3
Query Match:    98.94%      Indels:      0
DB:              6          Gaps:         0

US-09-985-689A-3 (1-433) x E03808 (1-2218)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnTyrGlyLeuTyr 20
Db 824 AATGATGTACAGAGGAGTAGTAAAGCTGATGTTGCCAACAACATACCGATTATAT 883

QY 21 GlyClnGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 884 GGCAAGGTCACATAGTTGAGTACGCGGACACAGGCTTAGATACAGGTCGTACAGTAGT 943

QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 944 TCTATGATGTAGACATTCGCGGGGAAATCACAGCTCTTTACCGTTAGAACAACTAAT 1003

QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1004 AATGCGAGTATCCGAATGGCGATGGCACATGATAGAGGTTCTGTCTTGGTAATGCT 1063

```

```

QY 81 LeuAsnTyrGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
Db 1064 TTTAATTAAGAAATGGCTCCGCAAGCTTAATCTTCTTCAATCTATTATGGTAGCAGC 1123

QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrIleuPheSerGlnAlaTrpAsnAla 120
Db 1124 GGAGGATTAGTGCTTACCATTGAACTTAATACAGCTTATTTAGTCAAGCTTGGAAATGCT 1183

QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyValProValAsnGlyAlaTyrThrAlaAsn 140
Db 1184 GGAGCAAGAATTCATATACTCTTTGGGAGGCCCCAGTAATAATGGAGCGTACATGCTAAC 1243

QY 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
Db 1244 TCGAGACAAGTCGATGAGTATGTTGAAATAATGATATACGCTACTTTTTCAGCTGGT 1303

QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1304 AATGAGGTCCTTAATTCAGGAACAATAGTGCTCCAGGTACACGGAATAATGCTATTACG 1363

QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 1364 GTGCGCGCAACGGAAACTATCGCCCAAGCTTCGGTTCGATAGACAGATAACCCAAATCAT 1423

QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 1424 ATTGCACAAATTCATCGAGAGGAGCTACGAGGAGTGGAGCAATTAAGCCTGACGTAACA 1483

QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 1484 GCTCTCGAACATTTATTTTATCAGCACGTTCTTCTTAGCTCCAGACTCTTCGTTTGG 1543

QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
Db 1544 GCGAATTAACACAGTAATAACGCTATATGCGGCGGCTACCTCCATGCGCACACCTATTGT 1603

QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
Db 1604 GCAGGGAATGTCGCGCAATTAACGTCGACATTTTATAAAATAGAGGTATTATCTCTAAG 1663

QY 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
Db 1664 CTTCTTTAATAAAGCTGCATTTATCGCTGGTGTACTGATGTTGGTTTAGGATATCCT 1723

QY 301 SerGlyAspGlnGlyTyrGlyArgValThrValTyrAspLysSerLeuAsnValAlaTyrVal 320
Db 1724 AGTGTGACCAAGCTCGGGCGGTGTACTCTAGATAAATCGTTAATGATGAGGTATGTC 1783

QY 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
Db 1784 AATGAAGCAACTGCATTTAGCCACAGGACAAAAGCAACGATTTCGTTCCAAGCACAAAGCG 1843

QY 341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
Db 1844 GGTAAACCTTTAAAAATCTCGTTAGTATGGACAGATGCTCCTGGAGTACAACTGCACT 1903

QY 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
Db 1904 TATACACTAGTTAATGATTAGATCTAGTTATTACTGCTCCGAATGACAAAATAATGTA 1963

QY 381 GlyAsnAspPheSerTyrProTyrAspAsnThrAspGlyArgAsnValGluAsn 400
Db 1964 GGAAATGATTTAGTTATCTCTTATGATAAATACCTGGGATGCTCGCAACAATGTTGAGAAC 2023

QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
Db 2024 GTATTATAACGCTCCGCAATCTGGAACTGATATAATTAATTCAGGTTCAAGCGTATAATGTA 2083

QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 2084 CCATCTGGCCACAGCGTTTCTCCTACTAGCTATCGTACAT 2122

```


RESULT 4
 AB046404
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 JOURNAL
 TITLE
 FEATURES
 source

AB046404
 Bacillus sp. Y PROC gene for protease, partial cds.
 AB046404.1 GI:12381940
 Bacillus sp. Y
 Bacillus sp. Y
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 (sites)
 Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
 Horikoshi, K.
 Novel oxidatively stable subtilisin-like serine proteases from
 alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 evolutionary relationships
 Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 20568675
 11118284
 2 (bases 1 to 1299)
 Saeki, K.
 Direct Submission
 Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
 Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
 Tel:81-285-68-7400, Fax:81-285-68-7403)
 Location/Qualifiers
 1..1299
 /organism="Bacillus sp. Y"
 /mol_type="genomic DNA"
 /strain="Y"
 /db_xref="taxon:133779"
 1..1299
 /gene="PROC"
 <!.>1299
 /gene="PROC"
 /codon_start=1
 /evidence="experimental"
 /transl_table=1
 /product="protease"
 /protein_id="BAB21267.1"
 /db_xref="GI:12381941"
 /translation="NDVARGIVKADVAQNNGYVGQGVVAVADTGLDGRNDSMHE
 AFRGITALYALNADSPNGHGTAVGSLGNALNKGMAPQANLVFOSIMDSGG
 LGLPSNLNTLPSQANWAGARIHTNSGAPVNGAYTANSROVDYVENVNMTVLFAAG
 NEGPSNGTISAPGTAKNAITVGATENVRPSPGSTADPNHIAQFSSFGATEDGRIPD
 VTAGTFLSRSLAPDSFWANYSKYAMGTSMTATPIVAGNVQLREHETKNG
 ITPRSLKALIAIGADIDVGIGYNGDQGWGRVTLNKSINVAIVNEATALATGKATY
 SFOAGAGPLKISLVMTDAPGSTTASVTLVNDLVLITAPNGQKYGVNDFSPYDNNW
 DGRNENVFINAPQSGTYIIIEVQAYNPSPQRFSLAIVH"

gene

CDS

Alignment Scores:
 Pred. No.: 2,1e-132 Length: 1299
 Score: 2234.00 Matches: 427
 Percent Similarity: 99.31% Conservative: 3
 Best Local Similarity: 98.61% Mismatches: 3
 Query Match: 98.72% Indels: 0
 DB: 1 Gaps: 0

ORIGIN

US-09-985-689A-3 (1-433) x AB046404 (1-1299)

Qy 1 AenAspValAlaArgGlyLeValValysAlaAspValAlaGlnAenAsnTyrGlyLeuTyr 20
 Db 1 AATGATGAGCAAGAGGATAGTAAAGCTGATGTTGCACAAAACAATTACGGATTATAT 60
 Qy 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 51 GCACCAAGGTCAAGTAGTTCGAGTAGCGACACAGGCTTAGATACAGGTCGTAACGATGT 120
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGATGAAGCATTCGCGGGGAAATCACAGCTCTTTACGCGTTAGGAAGAACTAAT 180

Qy 61 AenAlaAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAenAla 80
 Db 181 AATGCCAGTATCCGAATGGCATGGCACACATGTAGCAGGTTCTGTATTGTAATGCT 240
 Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAenLeuValPheGlnSerIleMetAspSerSer 100
 Db 241 TTAATAAAGGAATGCTCCGCAAGTAACCTAGTCTTCCAAATCTATTATGATAGCAGC 300
 Qy 101 GlyGlyLeuGlyGlyLeuProSerAenLeuAenThrLeuPheSerGlnAlaTrpAenAla 120
 Db 301 GGAGGATTTAGTGCTTACCATCGAACTTAAATACGTTATTATTAGTCAAGCTTGGAAATG 360
 Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrAlaAsn 140
 Db 361 GGAGCAAGAATTCATCTAATCTTGGGAGCCCAAGTAATCGAGCGTACACTGCTAAC 420
 Qy 141 SerArgGlnValAspGluTyrValArgAenAsnAspMetThrValLeuPheAlaGly 160
 Db 421 TCGAGACAAGTGGATGAATGTTCCGAAATAATGATATGACGGTACTTTTTGCAGCTGT 480
 Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 481 AATGAAGTCTTAATTCAGGAACAATTAGTGTCCAGGTACAGCGAAAAAATGCTATTACG 540
 Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
 Db 541 GTCGGCGCAACGAAAACTATCCCAAGCTTCGGTTCGATAGCAGATTAACCAATCAT 600
 Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 601 ATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCTCAGCTAAC 660
 Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 Db 661 GCTCCTGGAACTTTATTTTATCAGCACGTTCTTCTTAGCTCCAGACTCTTCGTTTGG 720
 Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
 Db 721 CGAATTATACAGTAATACCGTATATGGCGGTACCTCCATGCGGACACTATTGTT 780
 Qy 261 AlaGlyAsnValaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
 Db 781 GCAGGGAATGTCGCGCAATTAGCTGAGCAATTTATAAAAAATAGAGGTATTACTCTTAAG 840
 Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
 Db 841 CCTTCTTTAATAAAGCTGACTTATCGCTGCTGCTACTGATGTTGGTTTAGGATATCT 900
 Qy 301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValaLysVal 320
 Db 901 AATGTTGACCAAGGCTGGGGCGGTGTTACTCTAAATAAATCGTTAAATGTCGATATGTC 960
 Qy 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
 Db 961 AATGAAGCAACTGCATTAGCCACAGACAAAAGCAACGATTCGTTCACACCAAGG 1020
 Qy 341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
 Db 1021 GGTAAACCTTTAAATACTCGTTAGTAGCAGATGCTCCTCGGAAGTACAACTGCATCT 1080
 Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
 Db 1081 TATACACTAGTTAATGATTTAGTATCTAGTTATTACTGCTCCGAATGGCAAAAATATGTA 1140
 Qy 381 GlyAsnAspPheSerTyrProTyrAsnAenAsnTrpAspGlyArgAsnAsnValGluAsn 400
 Db 1141 GGAATGATTTTAGTTATCTTATGATAATACTGGGATGGTCGCAACATGTTGAGAAC 1200
 Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
 Db 1201 GTATTATAACGCTCCGCAATCTGCAACGTATATAATTAGGTTCAAGCGTATAATGTA 1260

QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
DB 1261 CCATCTGGCCACAGCGTTTCTCACTAGTATCGTACAT 1299

RESULT 5
LOCUS AR069954 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3003)
AUTHORS Sloma,A. and Christianson,L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES
1..3003
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1e-119 Length: 3003
Score: 2040.00 Matches: 385
Percent Similarity: 94.46% Conservative: 24
Best Local Similarity: 88.91% Mismatches: 24
Query Match: 90.15% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-3 (1-433) x AR069954 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValHisAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
DB 1470 AATGACGTGGCCGTTGGCATTTGAAAGCAGAGCTGCGACAAATAACTTTGGCTTATAT 1529

QY 21 GlyGlnGlyValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 1530 GGACAGGACAGATTTAGCAGTTGCTGATCTGAGCTTGCATGAGGAAAGAAATGACAGT 1589

QY 41 SerMetHisGluAlaPheAspGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 1590 TCGATGATGAGCATTCGCGGTAGATTTACCGCTATATGCTTGGCAGACGAT 1649

QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
DB 1650 AAGCCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGGAAATGCT 1709

QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
DB 1710 ACAATTAAGGATGGCCAGCCAGCAATCTAGTCTTCAATCTATTATGATAGTGT 1769

QY 101 GlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
DB 1770 GGAGGGCTGGAGGACACCTGCTAATCTACAAACATTTATTCAGTCAAGCATATAGTGT 1829

QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyValProValAsnGlyAlaTrpAlaAsn 140
DB 1830 GGACGAGATTCATACGATTTATGAGGAGGCTCCAGTAAACGGTCCCTATACGACAG 1889

QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
DB 1890 TCTCGAAATGTTGATGATTATGTGAGAAAAATGATATGACGATTTCTTTTGGCGCGGA 1949

QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB 1950 AATGAGGACAGGTAGCGGTACATCAGTGCACAGGAAACAGCAAAATGCGATTACA 2009

QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
DB 2010 GTTGGGCAACCGAAACCTACGCTTCCGATCTTATCGGATATATTAACCAT 2069

QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyValGlyLeuLysProAspValThr 220
DB 2070 GTTGCTCAATTCCTTCACAGGTCCTACTAGATGGAGCTATTAAGCGGAGCTCATG 2129

QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
DB 2130 GCACCAAGGTACGTATATTCCTCTGTAGATCATCATTAAGTCCAGATTCCTCATCTGG 2189

QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
DB 2190 GCAACCATGATAGTAATATGCTCATGGGTGCTACTTCTATGGCTACTCCATTTGTA 2249

QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
DB 2250 GCAGGTAACTTCACAAATTAAGGAGCATTTTGTGAAAAATAGAGGGTAACCTCCTAAG 2309

QY 281 ProSerLeuLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
DB 2310 CCTTCCCTTTTAAAGCTGCTTTAATTCAGGTGCTGCGATGTTGGACTTGGCTTTCCA 2369

QY 301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaLafyVal 320
DB 2370 AATGTTAACCAAGGATGGGAGAGTAACGTTAGATAAATCCCTAATATGTCGATTTGTG 2429

QY 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
DB 2430 AATGAAACGAGCCCTTTATCAACAGTCAAAAGCAACATATTCGTTTACGGCTCAAGCT 2489

QY 341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
DB 2490 GGTAACCCCTTAAATATATCATTGTTGTCAGATGCACAGGTAGCAGGCGCATCA 2549

QY 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
DB 2550 CTAATCTTAGTGAATGATTTAGACTTAGTAATCAGTGCACCAATAGAACTAAATACGTC 2609

QY 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
DB 2610 GGAATGACCTTACAGCACCGTATGATAACAATGGATGGCAGAACACACGTTGGAAT 2669

QY 401 ValPheLeuAsnAlaProGlnSerGlyThrTyrIleGluValGlnAlaLafyAsnVal 420
DB 2670 GTGTTTATCAATGCTCTCTCAAGCGGACGTATACAGTCGAAAGTGCAGGCTTCAATGTA 2729

QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
DB 2730 CCAGTAAGTCCGCAACCTTTCTTTAGCGATTTGATCAT 2768

RESULT 6
LOCUS BD062155 3003 bp DNA linear PAT 27-AUG-2002
DEFINITION Nucleic acids encoding a polypeptide having protease activity.
ACCESSION BD062155
VERSION BD062155.1 GI:22607760
KEYWORDS JP 2001514529-A/39.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3003)
AUTHORS Sloma,A. and Christianson,L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: JP 2001514529-A 39 11-SEP-2001;
NOVO NORDISK BIOTECH INC
COMMENT PN JP 2001514529-A/39
PD 11-SEP-2001
PF 09-JUN-1998 JP 1999503145
PR 12-JUN-1997 US 08/873479
PI ALAN SLOMA,LYNNE CHRISTIANSON
PC C12N15/57,C12N15/75,C12N9/54,C12K14/00
CC Strandedness: Single;
CC Topology: Linear;

FEATURES

source

EH Key

Location/Qualifiers.

Location/Qualifiers

1..3003
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 1e-119 Length: 3003
Score: 2040.00 Matches: 385
Percent Similarity: 94.46% Conservative: 24
Best Local Similarity: 88.91% Mismatches: 24
Query Match: 90.15% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-3 (1-433) x BD062155 (1-3003)

QY 1 AsnAspValAlaArgGlyLeuValValAlaGlnAsnAsnTyrGlyLeuTyr 20
DB 1470 AATGACGTGGCGCGTGGCATTTGTGAAGCAGACGCTCGCACAAATAAATTTGGCTTATAT 1529
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 1530 GCACAGGACAGATTGTAGCAGTTCTCTGATCTGCGCTTGTATACAGGAGAAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 1590 TCGATGTCATGAAGCATTCGCGGTGAAGATTACCGCACTATATGCACTGGGCGAGAACGAAT 1649
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
DB 1650 AACGCCAATGATCCAAATAGGACATGGAAACCCATGTTCTGATCTGTGTAGGAAATGCT 1709
QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
DB 1710 ACAATAAAGGATGGCAGCGGCAACCAATCTAGTCTTTTCAATCTATTATGATAGTGGT 1769
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrPheAsnAla 120
DB 1770 CGAGGCGCTGGGAGGACTACTCTCTAATCTACAAACATTTATTCAAGTCAAGCATATAGTGTCT 1829
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValAlaTyrThrAlaAsn 140
DB 1830 GAGCGGAGATTTCATACGAAATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACAGAC 1889
QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
DB 1890 TCTCGAAATGTTGATGATTATGTGAGAAAAAATGATGACGATTCTTTTTCGGCCCGGA 1949
QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB 1950 AATGAGGACCGAGTAGCGGTACATCATGTGACCGAGAACAGCAAAAATGCGATTACA 2009
QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
DB 2010 GTTGGGCAACCGAAACCTAGCTCCAGCTTCGGATCTTATGCGGATAATATTAAACCAT 2069
QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgGlyLysProAspValThr 220
DB 2070 GTTGCTCAATTCTCTACAGAGTCTCTAGATGATGACGATTAAAGCGGAGCTGATG 2129
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
DB 2130 GCACCAAGTACGTATATTCTCTCTGATAGATCATCATATTAGTCTCCAGATTCCTCATTTCTGG 2189
QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
DB 2190 GCAACCATGATAGTAAATATGCTCATGAGTGGTGTACTTCTATGGCTACTCCAAATTGTA 2249
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
DB 2250 GCAGGTAAATGTCACAAATTAGGGAGCATTTGTGAAAAATAGAGGGGTAACCTCTCTAAG 2309

QY 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
DB 2310 CTTTCCTTTTAAAGCTGCTTTTAAATTCAGGTGCTGGGATGTTGGACTTGGCTTTTCA 2369
QY 301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
DB 2370 AATGGTAACCAAGGATGGGAAGAGTAACGTTAGATAAAATCCCTAAATGTCGCAATTTGTG 2429
QY 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
DB 2430 AATGAACAGGCCCTTTTCAACAAGTCAAAAGCAACATATTCGTTTACGGCTCAAGCT 2489
QY 341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
DB 2490 GGTAAACCCCTTAAATAATATCACTTGTGGTCAGATGCACAGGTAGCAGCAGCGCATCA 2549
QY 361 TyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
DB 2550 CTAACCTTTTGTGAATGATTAGACTTAGTAATCACTGCACCAATGGAACTAAATACGTC 2609
QY 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
DB 2610 GGAATGACTTTACAGCACCGTATGATAACAATTTGGATGGCAGAAACAACCTGGAAAT 2669
QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
DB 2670 GTGTTTATCAATGCTCTCTCAACCGGACGTATACAGTCGAAGTGCAGGCTTACAAATGTA 2729
QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
DB 2730 CCAGTAAGTCCGAAACCTTTTCTTTAGCGATTGTACAT 2768

RESULT 7

AB046406
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AB046406 1302 bp DNA linear BCT 23-JAN-2001
Bacillus sp. NV1 PROE gene for protease, partial cds.

AB046406 GI:12381944

Bacillus sp. NV1

Bacillus sp. NV1

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

SAEKI, K., OKUDA, M., HATADA, Y., KOBAYASHI, T., ITO, S., TAKAMI, H. and

Horikoshi, K.

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic Bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

20569675

11118284

2 (bases 1 to 1302)

SAEKI, K.

Direct Submission

Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,

Tel:81-285-68-7400, Fax:81-285-68-7403)

Location/Qualifiers

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

FEATURES

source

gene

CDS

```
/db_xref="GI:12381945"
/translation="NDVARGIVKADVAQSYGLYGGQGVAVADTGLDTRNDSSMHE
AFRGKTAIYALGRNNDNGHCHTHVAGSVLNGTSGKMPAOLVFOQVNDVNSG
GLGGLPSVTLFQSAVAGARHTNSWGA PVNGAYTTDSRVDVYRKNDMVLPAA
GNEGNGGTISAPGTAKNAITVGTENLRPSFGYADINRHVAFSGRGTGGRIPK
DVMAPGTILSARSLADPSDFWANHDSKYAIMGTSNATPIVAGNVQALREHFKR
GITPKPSLKALAGATDILGLGYPNGQGRVTLDRKLNVAFNVTSLSTNQKAT
YSFTAQSGKPLKISLWSDAPASTASVTLVNDLDELVTAPNGTKYVGNDFTAIDYN
WDGRNNVENVFNAQSGTYTVEQAVNVPOGPOAFSLAIVN"

ORIGIN
Alignment Scores:
Pred. No.: 6,36e-119 Length: 1302
Score: 2020.50 Matches: 383
Percent Similarity: 94.70% Conservative: 28
Best Local Similarity: 88.25% Mismatches: 22
Query Match: 89.28% Indels: 1
DB: 1 Gaps: 1

US-09-985-689A-3 (1-433) x AB046406 (1-1302)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AACGATGTTGCCAGAGCATTGTAAAAGCCGATGTTGCCAGAGCAGCTATGGTTTATAT 60
Qy 21 GlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyValArgAsnAspSer 40
Db 61 GGCACAGGCGAAGTGTTCAGTAGCGGATACCGACTGGATGATACAGGCGGTAAATGACAGC 120
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyValArgThrAsn 60
Db 121 TCGATGCATGAACGCTTCGAGGAAAGATTACCGCGATATATGCGCTTGGAGAGCAAAAC 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyValAsnAla 80
Db 181 AACGCCAATGATCCAAACGACACGCGACGCGATGTTGCCGGAATCTGTTTAGAAGAACGCT 240
Qy 81 Leu---AsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 ACTTCGATAAAGGATGGCTCCACAAAGCTAATCTAGTTTTCATCTGTTATGGACAGC 300
Qy 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrIleuPheSerGlnAlaThrAsn 119
Db 301 AATGGCGGACTTCGCGGACTTCCTTCCAAATGTATGATACATTATTCAGCCAGGACATATAGT 360
Qy 120 AlaGlyValaArgIleHisThrAsnSerTrpGlyValaProValAsnGlyAlaTyrThrAla 139
Db 361 GCGGTGCCAGATCCATACAACTCATGGGAGCGCTGTAAACGAGCGTACACTACT 420
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCCAGAAACGTAGACGATTATGTTCTGTAATAAATGATATGCGCGTCTCTTTTTCAGCG 480
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 GGTAAACGAGGCGGAATGCGGACACATCAGTCTCTGTTACCGCGAGATGCTATC 540
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 541 ACAGTAGGGGCAACAGAAAACCTGCGCCCAAGCTTTGGATCTTTATGCTGACACATCAAT 600
Qy 200 HisIleAlaGlnPheSerSerArgGlyValaThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTACACAGTTTCTCTCCGCGGACCTTACAAAGGATGGAATCAACACCGGACGTA 660
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 661 ATGGCACCCAGGAACATTATTTATCGCAAGATCTTCTTGGCTCCGACTCCTCATTC 720
Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 721 TGGGCAAAACCATGATGACGAAATATGCTTATATGGGTGGAAACATCCATGGCGACACCGATT 780
```

AQSSGLVGGQGIIVAVATGTLDTGRDSSMHEAPRGKITALYALGRTRNANDTNGHGT
HVAGSLVNGSTKMGAPANLVFQSIIMDSGGGLGSLPSNLQTLFSAYGAGARIHTN
SLRPAFVAGATTTDSNRVDYVRKNDMTILFPAAGNEPNGGTISAPGTAKNAITVGATE
NKPFGGVADININHVAFSGRPTKDGRIKPDVNAPOGTFILSARSLAPDSFPAWH
DSKYAMGTGSMATPIVAGNVAQLREHFVKRGITPKPSLLKALJAGAADIIGLYEN
GNQGRVTLDPKSLNVAVNESSLSITSOXATYVATAGKPLKILISVWSDAPASTTA
SVTLVNDLULVITAPNGTQYVGNDFTSFINNWDGRNVENVFINAQSPQSYTILEVOA
INVPVGPQTFSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 2,13e-117 Length: 1923
Score: 1999.50 Matches: 381
Percent Similarity: 94.01% Conservative: 27
Best Local Similarity: 87.73% Mismatches: 25
Query Match: 88.36% Indels: 1
DB: 1 Gaps: 1

US-09-985-689A-3 (1-433) x AB084155 (1-1923)

Qy 1 AsnAspValAlaArgGlyValLeuValAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTTCACGCGGAAATGTCAAAGCGGATGTCGCGCAGCAGCAGCTCGGTTGTAT 678
Qy 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGCACAGGACACATCGTAGCGGTCCCGNATACAGGCTTGATACAGGTTCGCAATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGTCATGAAGCCTCCGGGGGAAATATCTGCATTATATGCAATGCGGACGCAAT 798
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATGATACGATGGTCATGTCAGCATGTGGCTCGCTCGATTATAGGAACGCG 858
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAAGGAATCGCGCTCAGGCGAATCTAGTCTTCCATCTATCATGATGATG 918
Qy 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
Db 919 GGTGGGGNCTTGGAGGACTACTTCGAACTGCGAACTTATTCAGCAAGCATACAGT 978
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValAlaTyrThra 139
Db 979 GCTGTGCGCAGAAATCATACAAATCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAla 159
Db 1039 GATTCAGAAATGTGGATGATGATGTGCGCAAAATGATATGACGATCCTTTTCGCTGCC 1098
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GCGAATGAAGGACCGAAGCGGGAACCATCATGTCAGTCACCGCAGCAGCTAAATGCAATA 1158
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACAGTCGGAGCTACCGAAACCTCCGCCCAAGCTTCGGGTCTTATGCGGCAATATCAAC 1218
Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGCGCACATCTCTTCAGTCGTGGACCGCAGCAAGAGATGGACGGATCAACCGGATGC 1278
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCACCAGGACGCTTATATATCAGCAAGATCTCTCTTGGACCGGATTCCTCTTC 1338
Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1339 TGGGCGAACCATGACAGTAAATATGATATCATGCTGGTGGAACTGCTCATCGCTACCCGATC 1398
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279

Db 1399 GTTGTCTGGAACGTGCGACAGCTTCGTGAGCATTTTGTGAAAACAGAGCGCATCACACA 1458
Qy 280 LysProSerLeuIleLysAlaLeuIleAlaGlyValThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCTTCTCTATTAAAGCGGCACTGATTCGGGTGCGAGCAGACATCGGCTTGGCTAC 1518
Qy 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGACCGTAAACCAAGGATGGGACGAGTGACATTGGATAAATCCCTAAACCTTGCTAT 1578
Qy 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1579 GTGACAGTCCAGTCTCTATCCACCCAGCAAAAGACGACGTACTCTGTTTACTGCTACT 1638
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1639 GCGGCAAGCCTTTGAAGATCTCCCTGGTATGCTGTGATGCCCTCGCAGCAACATGCT 1698
Qy 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCGTAAACGCTTGTCAATGACCTGGACCTTGCTATTCACCGTCCAAATGGCACATAT 1758
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1759 GTTGGAAATGACTTTACTTCGCCATACATAATACTGGATGGCGCAATTAACGTAGAA 1818
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGTATTATTAAATCGCCACAAAGCGGAGCGTATACAAATGAGGTACAGGCTTTAAAC 1878
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGTTCGACACAGACCTTCTCGTTGGCAATTGTGAAC 1920
RESULT 9
AB046403
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
gene
CDS

AB046403 1920 bp DNA linear BCT 10-MAY-2002
Bacillus sp. 9860 PROA gene for protease, complete cds.
AB046403 GI:205211152
Bacillus sp. 9860
Bacillus sp. 9860
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1
Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
Horikoshi, K.
Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
20568675
11118284
2 (bases 1 to 1920)
Saeki, K.
Direct Submission
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail: 387185@kasanet.kao.co.jp,
Tel: 81-285-68-7400, Fax: 81-285-68-7403)
On May 9, 2002 this sequence version replaced gi:12381938.
Location/Qualifiers
1. .1920
/organism="Bacillus sp. 9860"
/mol_type="genomic DNA"
/strain="9860"
/db_xref="taxon:133778"
1. .1920
/gene="PROA"
1. .1920
/gene="PROA"
/codon_start=1

/transl_table=11		1276		ATGGCGCCAGGACATACATTTTATCAGCAAGATCTCTCTTGACCCGATCTCTCTTC	1335
/product="protease"		240		TTTAAAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle	259
/protein_id="BAB21266.2"		1336		TGGCGGAATCATCAGCAAGAAATGCTATATGGTGGGAACGCTCCATGGCAACCGAT	1395
/db_xref="GI:20521153"		260		ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLeuAsnArgGlyIleThrPro	279
/translation="WRKKVFLSVLSAAALSTVALNPNPSAGDTFDLDPKGIQTTF		1396		GTTCGGGGGAATGTTCACAGCTCCGTGAGCATTTTGTGAAAAATAGAGAAATCACTCCT	1455
DVSFGSKQRTGAAFLVSESNVLLKLLKLETPANNKLHIVQFNPLBETKOK		280		LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr	299
LTKALADLPKNNKSVQLRGIEIAQYASNDVHVIITAKPEYKMDVARGIVKADVA		1456		ARGCCTTCCTATTTGAAGACAGCTTTGATTGCAAGGTGCTGCTGATTTGGATTGGATTAT	1515
QSSYGLYGCGQVAVADTGLDGRNDSMHEAFRGKITALVALGRTNNANDTNGHGT		300		ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr	319
VAGSVLGNATKNGAPQANLVFQSIMDSGGGLGPNLQTLFSQAFSAGARLHTNS		1516		CCGAACGAAACCAAGGATGGGCGCGAGTGCCTGGATAAATCGTTGAACGTTGCCTAT	1575
WGAAVNGATTTDSNRVDDIVRKNDITLFAAGNPNPGTISAPGTAAGATLVGATEN		320		ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln	339
LRPSFGSADNTHNVAQSFSSRGPTDKRIKPDWAPGTIILSARSSLAPDSSFWANHG		1576		GTGAACGATCCAGTGGCCCTTATCAACTAGCCAAAAGCGACATATACCTTTACTGCAACG	1635
SKYAYNGTSMATPIVAGNVQREHFVKNRGITKPSLLKALITAGADYGLYPNG		340		AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrAla	359
NOGWGRVTDKSLINAVYNESALSTSKATYTFATAGKPLKISLVMSDAPASTAS		1636		CGCGGCAAGCAATTTGAATAATCTCCCTGGTATGGTGGATGCCCTGCAACGACTACTGCT	1695
VTILNDLIVITAPNGTRVYVGNDFSAPEFNWNGNENVFNINSQSGTITIEVQAY		360		SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr	379
NVPVGPQNFSLAVN"		1696		TCTGTAAACCTGGTCAATGATTTGGATTGGTCAATACAGCAACCAACGAAAGATAT	1755
ORIGIN		380		ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGlu	399
Alignment Scores:		1756		GTGCGGAATGACTTCTCAGCACCATTTGACAATACTGGCATGGCCGCAATAACGTAGAA	1815
Pred. No.:		400		AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn	419
Score:		1816		AATGTATTATTAAATTCGCCCAAGTGGAACTATATACCATTTGAGTGGCAAGCATATAAT	1875
Percent Similarity:		420		ValProSerGlyProGlnArgPheSerLeuAlaIleValHis	433
Best Local Similarity:		1876		GTGCGGTGGACCAAAACTTCTCGTTGGCAATTGTGAAC	1917
Query Match:		RESULT 10			
DB:		AR368116		1920 bp	
US-09-985-689A-3 (1-433) x AB046403 (1-1920)		LOCUS		DNA	
		AR368116		linear	
		DEFINITION		Sequence 3 from patent US 6376227.	
		ACCESSION		AR368116	
		VERSION		AR368116.1	
		KEYWORDS		GI:34601777	
		SOURCE		Unknown.	
		ORGANISM		Unknown.	
		REFERENCE		Unclassified.	
		AUTHORS		1 (bases 1 to 1920)	
		TITLE		Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,	
		JOURNAL		Shikata, S. and Nomura, M.	
		FEATURES		Alkaline protease	
		source		Patent: US 6376227-A 3 23-APR-2002;	
				Location/Qualifiers	
				1..1920	
				/organism="unknown"	
				/mol_type="genomic DNA"	
		ORIGIN			
		Alignment Scores:			
		Pred. No.:		2,46e-117	
		Score:		1998.50	
		Percent Similarity:		94.01%	
		Best Local Similarity:		88.02%	
		Query Match:		88.31%	
		DB:		1	
		Length:		1920	
		Matches:		382	
		Conservative:		26	
		Mismatch:		25	
		Indels:		1	
		Gaps:		6	
		US-09-985-689A-3 (1-433) x AR368116 (1-1920)			
		1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr		20	
		616 AATGATGTGCCAGAGGATTTGTCAAAGCGGATGTGGCAGACGACTACGGTTTGAT		675	
		21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer		40	
		676 GGCAAGGCCAGATTTGCGAGTTGCGATCTACTGGATTGGATACAGAAAGAAACGACAGT		735	
		41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn		60	
		736 TCGATGATGAGGCTTCGCGGTGAAATAACAGCAGCTATATGCATGGGTGGAGCAT		795	
		61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn---		79	
		796 AATCGGAATGATACGAACGGTTCATGGTACCATCTGCGAGTTTCGGTATTAGAAATGCG		855	
		80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer		99	
		856 GCAACGAATAAGGAATGGCACCTCAAGCGAATCTGTTTTCATTCATCATCTGGATAGC		915	
		100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaIleTyrAsn		119	
		916 AGTGTGGCTTGGAGGCTTCCCTTCAATCTGCAACCTTATTCAGCCCAAGCATTCAGT		975	
		120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla		139	
		976 GCAGGTGCCAGAAATTCATACAACTCTCTGGGGGCGACGGTGAATGGGCTTACAGACA		1035	
		140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla		159	
		1036 GATTCAGAAATGTGGATGACTATGTAAGGAAATATGATGACGATCTCTTTTCGGCGCT		1095	
		160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle		179	
		1096 GGGGAATGAAAGCCGACGGCGGTACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		1155	
		180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn		199	
		1156 ACAGTGGCGGACCGAANAACCTGGTCCAGCTTCGGTTCCTATGAGATATATTAAC		1215	
		200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal		219	
		1216 CACGTTGCACAGTCTCTTCCCGTGGCCGCAAAAGATGGCGGAATCAAGCCTGATGTC		1275	
		220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPhe		239	

QY 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGGTATTGTCAAAGCGGATGTGGCAGACAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyValGlnAsnAspSer 40
Db 676 GGACNAGGCCAGATTGTCCGAGTTGCCGATCTGGATTGGATACAGGAAGAACGACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAAGCCTTCGCGGGTAAATAACAGCAGCTATTATGCACCTGGGTGCGACGAAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATGCGAATGATACGAACGGTCATGTTACCATGTGGCAGTTCGGTATTAGGAATGGC 855
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGAATAAGGAATGGACCTCAAGCGAATCTGGTTTTTCAATCCATCATGGATAGC 915
QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 916 AGTGTGGGCTTGGAGGCTTGCTTCCCAATCTGCAAACTTATTTCAGCCAGCAATTCAGT 975
QY 120 AlaGlyValaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyValaTyrThrAla 139
Db 976 GCAGTGGCCAGATTATCAACAACCTCTGGGGGGCAGCGGTGAATGGGGCTCACGACA 1035
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1036 GATTCAGAAATGTGGATGATATTAAGGAAATGATATGACGANTCTTTTCGCGCT 1095
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1096 GGGATGAAGCCGACCGCGGTACCATCATGTGCACCTGTGTACGGCTAATAAGCCATA 1155
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1156 ACAGTCGGCGCAACCGAAACCTGGTCCAGCTTCGTTCCGTATGCAGATAATATTAAAC 1215
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyValGlyIleValProAspVal 219
Db 1216 CAGGTGACAGTTCTTCCGTGGCGGACAAAGATGGCGGAATCAACCTGATGTC 1275
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAGATCTTCTTTCACCCGATTCTCTCCCTC 1335
QY 240 TrpAlaAsnTyrAsnSerIlystYrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 1336 TGGGCGAATCATGACAGCAAAATATGCTTATATGGGTGGAACGTCCATGGCAACCCGATT 1395
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1396 GTTGGGGGAATGTTGCACAGCTCGGTGAGCATTGTTGMAAAATAGAGGAATCATCTCT 1455
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1456 AAGCCTTCCCTATTGAAGCAGCTTGTATGTGAGTGTCTGTGATGTTGGATTGGGTTAT 1515
QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1516 CCGAACGGAACCAAGGATGGGGCGAGTGACCTGGATTAATCTGTGAACGTTGCCCTAT 1575
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1576 GTGAACGAATCCAGTCCCTATCACTAGCCAAAAGCGACATATACCTTACTGCAACG 1635
QY 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
Db 1636 GCGGGCAGCCATTGAATATCTCCCTGGTATGGTGGATGCCCCCTGCAGACACTACTGCT 1695

QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1696 TCTGTAAACCTCGTCAATGATTGGATTGGTCATTACAGCACCACCAACGAAACAGATAT 1755
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1756 GTCCGGAATGACTTCTCAGCACCATTTGACATACTGGGATGGCGCAATAACGTAGAA 1815
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1816 AATGTATTATTAAATTCGCCCAAGTGAACATATACCATTSAGGTGCAAGCATATAAT 1875
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1876 GTGCCGTTGGACCAAAAACCTTCTGTTGGCAATTGTGAAC 1917
RESULT 11
LOCUS AR368118 1923 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 7 from patent US 6376227.
ACCESSION AR368118
VERSION AR368118.1 GI:34601779
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1923)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S. and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 7 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1923
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3,82e-117 Length: 1923
Score: 1995.50 Matches: 380
Percent Similarity: 94.01% Conservative: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.18% Indels: 1
Gaps: 6
US-09-985-689A-3 (1-433) x AR368118 (1-1923)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTTCGCGCTGGAAATTTGTCAAAGCGGATGTGCTCAGAGCAGCTACGGTTGTAT 678
QY 21 GlyGlnGlyGlnValLalAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGTCGCAATGACGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGCATGAACCTTCCCGGGAAATTTACTGCATTATATGATTGGACGACGAT 798
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATCATACGAATGTGTCATGTCATGCGATGTCGCTCGCTCCGTATTAGGAAACGGC 858
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTTAATAAGAAATGGCGCTTCAGCGAATCTAGTCTTCCAACTCATCATGGATAGC 918
QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 919 GGTGGGGGACTTGGAGGACTTACCTTCGAATCTGCAACCTTATTCAGCAAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyValaTyrThrAla 139

Db 979 GCTGTGCGCAGATTTCATACAACTCTGCGGAGCAGCAGTGAATGGGCTTACACACA 1038
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATATGATACGACCTCTTTGCGTCC 1098
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIleAsnAlaIle 179
Db 1099 GGAATGAAGGACCGAACCGCGGACCAATCATGTGACCGGACAGCTTAATAATGCAATA 1158
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACAGTCGGAGCTACGGAAACCTCGCCCAAGCTTTGGGTCTTATGCGGCAATATCAAC 1218
Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleIleProAsnVal 219
Db 1219 CATGTGGCAGTTCCTTCACGTGGACCGACAAAGATGGACGGATCAACCGGATGTC 1278
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 1279 ATGGCACCGGAACGTTCTACTATCATCATGACAGATCTCTCTTGCACCGATTCTCTTC 1338
Qy 240 TrpAlaAsnTyrAsnSerIleTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGAACCATGACATGATAATGATCATGATGCGGTGGAACTCCATGGCTACACCGATC 1398
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1399 GTTGCTGGAAACGTGGCAGCTTCGTGAGCATTTTGTGAAAAACAGAGGATCACACCA 1458
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCCTTCTCTATTAAAGCGGCACCTGATTGCGGTGACAGCTGACATCGGCTTGGCTAC 1518
Qy 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGACGGTACCAAGATGGGACGATGACATTTGGATTAATCCTGACGTTCCCTAT 1578
Qy 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1579 GTGACGAGTCCAGTCTCTATCCACCGACCAAAAGCGACGTACTCGTTACTGCTACT 1638
Qy 340 AlaGlySerProLeuIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
Db 1639 GCGCGAGAGCTTTGAAATCTCCCTGGTATGCTGATGCGCTCCCAATGGCACACAGTAT 1698
Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCGTAACGCTTGCAATGATCTGGACCTGTCTATACCGCTCCCAATGGCACACAGTAT 1758
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGlu 399
Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATGATGATGATGATGATGATGATGATGAT 1818
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGTAATTAATATGACCAACAAAGCGGACGCTATACAAATGAAGTACAGGCTTATAAC 1878
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGGTGGACACAGAACTTCTCGTGGCAATTGTGAAT 1920
RESULT 12
AX839476
LOCUS AX839476 1305 bp DNA linear PAT 15-DEC-2003
DEFINITION Sequence 2 from Patent EP1347044.
ACCESSION AX839476
VERSION AX839476.1 GI:39922766
KEYWORDS
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacillus sp. KSM-KP43
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1

AUTHORS Okuda, M.K., Sato, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K.,
Saeki, K.K., Kobayashi, T.K. and Nomura, M.K.
TITLE Alkaline protease
JOURNAL Patent: EP 1347044-A 2 24-SEP-2003;
Kao Corporation (JP)
FEATURES
source Location/Qualifiers
1. .1305
/organism="Bacillus sp. KSM-KP43"
/mol_type="unassigned DNA"
/db_xref="taxon:103322"
CDS
1. .1305
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAB85146.1"
/db_xref="GI:39322767"
/translation="NDVARGIVKADVAOSSYGLYGGQIVAVADTGLDTRNDSSMHE
AFGKITALYALGRTNNNDTNGHGTHTVAGSVLGGSTNKGMAPQANLVPQSIMDSGG
GUGLPSNIQTLPFSQISAGARIHTNSWGAANGAYTTDSRVNDDTFRKNDMLIFAA
GNEPGNGTISAPGTAKNAITVTGATENLSPFGSYADNINHVAGQSSGPTKDGRIKP
DVNAPGTFLSARSLAPDSFWANHDSKYVGGTSMATPIVAVAGSLRHFVKNR
GITPKSLKALIAAGADIGLYPNGNCGWRVTLDKSLNVAIVYNESSLSSTQKAT
YSTATAGKPLKISLVMSDAPASTTASVTLVDLIVITAPNGTQYVGNDFTSYNDN
WDRNNVNFVINAPOSQYTIIEVQAYNVPGPTFSLAIVN"

ORIGIN

Alignment Scores:
Pred. No.: 2,79e-117 Length: 1305
Score: 1994.50 Matches: 380
Percent Similarity: 94.01% Conservative: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.14% Indels: 1
DB: Gaps: 1

US-09-985-689A-3 (1-433) x AX839476 (1-1305)

Qy 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AATGATTTTGGCGTGGATTTCAAGCGGATGTGGCTCAGACGAGCTACGGGTTGTAT 60
Qy 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GCACAGACACAGATCGTAGCGGTTGCCGATACAGGCGTTGATACAGCTCGCAATGACAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyIleValLeuThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGATGAGGCTTCCCGGGGAAATTTACTGCAATATATGATGATGATGATGATGAT 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 181 AATGCCAATGATACGATGCTCATGTGACGATGCTGCTCGCTCGCTCGCTCGCTCGCT 240
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 TCCACTAATAAGGAATGGCGCTCAGCGGAACTTAGCTCTCCAACTTATCATGATGATGAT 300
Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
Db 301 GGTGGGGGACTTGGAGGACTACCTTCGAATCTGCAACCTTATTTCAGCCCAAGCATACAGT 360
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 361 GCTGTGCGCAATTCATAAACTCTCGGGGACGACGATGATGATGATGATGATGATGATGAT 420
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGATGATGATGATGATGATGATGAT 480
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 GGGAAATGAAGACCGCAACCGCGGAACCATCATGATGATGATGATGATGATGATGATGATGAT 540
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199

Db	541	ACAGTCGGAGTCACGGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGCAATATCAAC	600
Qy	200	HisIleAlaGlaPheSerArgGlyAlaThrArgAspGlyArgIleIleYsProAspVal	219
Db	601	CATGTGGCACAGTCTCTTCAGTGGACCGCAAAAGGATGGACGATCAAAACCGGATGTC	660
Qy	220	ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe	239
Db	661	ATGGCACCGGAACGTTTATCTATCATCAGCAAGATCTCTTGGACCGATTCCTCCTTC	720
Qy	240	TrpAlaAsnTyrAsnSerIleYsAlaThrMetGlyGlyThrSerMetAlaThrProIle	259
Db	721	TGGCGCAACCATGACAGTAAATATGATCATGATGGTGGAACTCCATGCTACACCGATC	780
Qy	260	ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleIleYsAsnArgGlyIleThrPro	279
Db	781	GTTGCTGGAAACGTGGCAGAGCTTGTGAGCATTTTGTGAAAAACAGAGGATCACACCA	840
Qy	280	LysProSerLeuIleYsAlaIleAlaIleAlaGlyAlaThrAspValGlyLeuGlyTyr	299
Db	841	AAGCCTCTCTATTAAAGCGCAGCTGATTCGCGGTGCAGCTGACATCGGCTTGCGCTAC	900
Qy	300	ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr	319
Db	901	CCGAACGGTAACCAAGATGGGAGGAGTGCATTCGATAAATCCTGACGTTGCCCTAT	960
Qy	320	ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln	339
Db	961	GTGAACGAGTCCAGTCTCTATCCACAGCCAAAGAGCAGCTACTCGTTTACTGCTACT	1020
Qy	340	AlaGlyLysProLeuYsIleSerLeuValTrpThrAspAlaProGlySerThrThrAla	359
Db	1021	GCGCGCAAGCCTTTGAAATCTCCCTGGTATGCTGATGCTGATGCTGCGGACCAACTGCT	1080
Qy	360	SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnYsTyr	379
Db	1081	TCGTTAACGCTGTCAATGATCTGACCTTGTCAATACCGCTCCAAATGGCACACAGTAT	1140
Qy	380	ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnValGlu	399
Db	1141	GTAGGAATAGCTTTACTTCGGCATACATGATGATCTGGATGGCGGCGCAATACGTAGAA	1200
Qy	400	AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn	419
Db	1201	AATGTATTATTAAATSCACCAAGCGGAGTATACATTAATGAGGTACAGGCTTATTAAC	1260
Qy	420	ValProSerGlyProGlnArgPheSerLeuAlaIleValHis	433
Db	1261	GTACCGGTTGGACACAGACCTTCTCGTTGGCAATTGTGAAT	1302
RESULT 13			
AB051423			
LOCUS			
AB051423			
DEFINITION			
Bacillus sp. KP43 PROF gene for protease, complete cds.			
ACCESSION			
AB051423.2			
VERSION			
AB051423.2 GI:20521154			
KEYWORDS			
Bacillus sp. KSM-KP43			
SOURCE			
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
ORGANISM			
Bacillus sp. KSM-KP43			
REFERENCE			
1			
Itoh, S. and Saeki, K.			
AUTHORS			
TITLE			
new protease			
JOURNAL			
Unpublished			
REFERENCE			
2 (bases 1 to 1923)			
AUTHORS			
Saeki, K.			
TITLE			
Direct Submission			
JOURNAL			
Submitted (21-NOV-2000) Kateuhisa Saeki, KAO, CORPORATION;			
2606, AKABANE, ICHIKAWA, HAGA, TOCHIGI 321-3486, Japan			
(E-mail: 38718@kasanet.kao.co.jp, Tel: 81285687471 (ex. 7471),			
Fax: 81285687403)			
COMMENT			
On May 9, 2002 this sequence version replaced gi:14164344.			

FEATURES		Location/Qualifiers	
source		1..1923	
		/organism="Bacillus sp. KSM-KP43"	
		/mol_type="genomic DNA"	
		/strain="KP43"	
		/db_xref="taxon:109322"	
gene		1..1923	
		/genes="PROF"	
CDS		1..1923	
		/genes="PROF"	
		/codon_start=1	
		/transl_table=11	
		/product="protease"	
		/protein_id="BAB55674.2"	
		/db_xref="GI:20521155"	
		/translation="MRKKKVFSLVLSAAAILSTVALSNPSAGGARNFLDFKGIQTT TDAGSKOGOTGAAPLVESENVLKPKLOKLETVPANLKLHIIOFNGPILEETKQ OLEKTKGKILDIYIPDYAVIVEGVKTSSTIEHVESVEPLYPIRIDPOLFTKGAS ELKVALNTPKQKKEVQIRGEIQAFALNDVLYITAKPEYKWNDAVARGIVADV AQSSYGLYGQGIQIVAVADTGLDTRGNSDHEAFRGTITATLALGRNTNANDINGHT HVASVILGGSTNKGAPQANLVFSIMDSGGGLGSPNLQTLFSQAYSAGARHNT SWGAANGAYTDSRVDDYVRKNDMTILFAAGNEGPGGTISAPGTAKNAITVGTAE NLKPSVSGADNINHVAFQFSGPTKDGRIKPDVMAFGPTILSARSLAPDSSEFWAH DSKVYMGGTSMATPIVAGNVAOLREHFVNKRGITPKPSLLKALLAGAADI GLGYPN GNQWGRVTLDKSLNVAYNVESLSSTOKATYSETATAGKPKIKISLVKSDAPASITA SVTLNLDLIVITAPNGTQYVGNDFTSYNDNWDGRNVNFINAPQSGTYTILEVQA INVPVGPQTFLSAIVN"	
ORIGIN		4,42e-117	
Alignment Scores:		Length:	
Pred. No.:		1923	
Score:		Matches:	
Percent Similarity:		380	
Best Local Similarity:		Conservative:	
Query Match:		28	
DB:		Mismatch:	
		25	
		Indels:	
		1	
		Gaps:	
		1	
		US-09-985-689A-3 (1-433) x AB051423 (1-1923)	
Qy	1	AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr	20
Db	619	AATGATGTTCCGCGTGAATTCGAAGCGGATGTGGCTCAGACGACGATCGGTTGTAT	678
Qy	21	GlyGlnGlyValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40
Db	679	GGACAAGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGTCGCAATGACAGT	738
Qy	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60
Db	739	TCGATGTCATGAACCTTCGCGGAAATTTACTGCAATATATGATTCATGGACGACGAAT	798
Qy	61	AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn---	79
Db	799	AATGCCAATGATACGAATGCTCATGTACGATGTGGCTGGCTCCGCTATTAGGAACGGC	858
Qy	80	AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	99
Db	859	TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAAATCTATCATGGATAGC	918
Qy	100	SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn	119
Db	919	GGTGGGGGACCTTGGAGGACTACCTTCGATCTGCAACCTTATTTCAGCCAGCATACAGT	978
Qy	120	AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla	139
Db	979	GCTGGTGCCAGAAATTCATCAAACTCCTGGGAGCAGCAGTGAATGGGCTTACACAACA	1038
Qy	140	AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla	159
Db	1039	GATTCCCAAGAAATGTGGATGACTATGCGCAAAAATGATATGACGATCTTTTCGTGCC	1098
Qy	160	GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	179

Db 1099 GGGAAATGAAGGACCGGAAACCATCATGTCACACAGGCACAGCTAAATAATCAATA 1158
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGACAATATCAAC 1218
Qy 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleValysProAspVal 219
Db 1219 CATGTGGCAGATTCCTTCACGTGGACCGACAAAGATGGAGGATCAACCGGATGTC 1278
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCACCGGAAACGTTCTATCATCAGCAAGATCTTCTTGCACCGGATTCCTCCCTTC 1338
Qy 240 TrpAlaAsnTyrAsnSerIleValAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGAACCATCATAGTAATATGATATCATGGTGGAACTCCATGGCTACACCGATC 1398
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleValysAsnArgGlyIleThrPro 279
Db 1399 GTTGTCTGAAACCGTGGCACAGCTTCGTGAGCATTTTGTGAAACACAGAGGCATCACCA 1458
Qy 280 LysProSerLeuIleValysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCCTTCTCTATTAAGCGGCACTGATTCGCCGTGCGAGCTGACATCGCCCTTGGCTAC 1518
Qy 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGAACGTTAAACCAAGATGGGACGAGTGACATTCGATAAATCCCTGAACGTGGCTAT 1578
Qy 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnIleValAlaThrThrSerPheGlnThrGln 339
Db 1579 GTGAACGAGTCCAGTCTCTATCCACAGCCAAAGCGGACGACTCTGTTTACTGCTACT 1638
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
Db 1639 GCCGGCAAGCTTTGAAATCTCCCTGTTATGTCGTGATGTCGCCCTCGAGCACCACTGCT 1698
Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCCGTAACTGTTGCAATGATCTGGACCTTGTATTACCGCTCCAAATGGCACACAGTAT 1758
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGlu 399
Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATACTGGATGGCCGCAATACGTAGAA 1818
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGTATTTATTAATGCACCAACGAGGACCGTATACAAATTGAGGTACAGGCTTATAAC 1878
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGGTTGGACCAACAGACCTTCTCGTTGGCAATGTGTAAT 1920

RESULT 14
AR368117
LOCUS 1923 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 5 from patent US 6376227.
ACCESSION AR368117
VERSION AR368117.1 GI:34601778

KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1923)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S., and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 5 23-APR-2002;
FEATURES Location/Qualifiers
1..1923
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores: 4.42e-117 Length: 1923
Pred. No.: 1994.50 Matches: 380
Score: 94.01% Conservative: 28
Percent Similarity: 87.56% Mismatches: 25
Best Local Similarity: 88.14% Indels: 1
Query Match: 6 Gaps: 1
DB: 6

US-09-985-689A-3 (1-433) x AR368117 (1-1923)

Qy 1 AsnAspValAlaAlaArgGlyIleValysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTGCGGTGAAATTTGCAAGCGGATGTGCTCAGAGCAGCTACGGGTTGTAT 678
Qy 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAGACAGATCTGAGGTTGCCATACAGGCTTGATACAGGTCCCATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGATGAAGCTTCCGCGGAAATTTACTGCATTATATGATTGGGACGACCAAT 798
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATGATACGAATGTCATGGTACGATGTGCTCGCTCCGTTATTAGGAACGGC 858
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAGGAATGGGCGCTCAGCGCAATCTAGCTTCCAATCTATCATGATGATG 918
Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
Db 919 GGTGGGACCTTGAGAGCTACCTTCGAATCTGCAACCTTATTACGCCAAGCATACAGT 978
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 979 GCTGGTGCAGAAATCATACAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
Qy 140 AsnSerArgGlnValAlaAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCCGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGGAAATGAAGACCGAACCGCGGAAACCATCATGTCACACGACAGCTAAATATGCAATA 1158
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGCAATATCAAC 1218
Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGGCAGATTCCTTCTTCAGTGGACCGCAAGGATGGACCGGATCAACCGGATGTC 1278
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCACCGGAAACGTTCTATCATCAGCAAGATCTTCTTGCACCGGATTCCTCCCTTC 1338
Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGAACCATCATAGTAATATGATATCATGGTGGAACTCCATGGCTACACCGATC 1398
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleValysAsnArgGlyIleThrPro 279
Db 1399 GTTGTCTGAAACCGTGGCACAGCTTCGTGAGCATTTTGTGAAACACAGAGGCATCACCA 1458
Qy 280 LysProSerLeuIleValysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCCTTCTCTATTAAGCGGCACTGATTCGCCGTGCGAGCTGACATCGCCCTTGGCTAC 1518
Qy 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319

Search completed: April 4, 2004, 08:06:52
Job time : 3595.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus p2n model

Run on: April 3, 2004, 16:37:54 ; Search time 335.908 Seconds
(without alignments)

5476.111 Million cell updates/sec

Title: US-09-985-689A-3

Perfect score: 2263

Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVPSGQRFSLAIVH 433

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus.p2n.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09985689/runat_31032004_161806_4145/app_query.fasta_1_3498
-DS=N Geneseq 29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CGN1_1_1238 @runat_31032004_161806_4145 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04: *
1: geneseqn1980s: *
2: geneseqn1990s: *
3: geneseqn2000s: *
4: geneseqn2001as: *
5: geneseqn2001bs: *
6: geneseqn2002as: *
7: geneseqn2003as: *
8: geneseqn2003bs: *
9: geneseqn2003cs: *
10: geneseqn2004as: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2239	98.9	1299	2	AAQ27516 Alkali-pr
2	2040	90.1	3003	2	AAV82382 Bacillus
3	1988	88.3	1920	2	AAx37277 Bacillus
4	1935.5	88.2	1923	2	AAx37279 Bacillus
5	1994.5	88.1	1923	2	AAx37278 Bacillus
6	443	19.6	1977	2	AAx37278 Bacillus
7	443	19.6	1977	2	AAx37278 Bacillus
8	422.5	18.7	1236	2	AAx05926 WO9856926
					AAx05920 Hyperther

9	422.5	18.7	1566	2	AAx85668 Pyrococcus
10	422.5	18.7	1962	2	AAx85668 Pyrococcus
11	422.5	18.7	1962	2	AAx05929 Hyperther
12	401.5	17.7	1977	2	AAx85669 Protease
13	370.5	16.4	2121	4	ABL54900 T. yonsei
14	346.5	15.3	135838	7	ABX34289 S. atrool
15	343.5	15.2	2539	2	AAx1454 Streptomy
16	343.5	15.2	2809	2	AAx1455 DhpA-mel
17	336	14.8	1329	6	ABK74643 Bacillus
18	312	13.8	2532	2	AAQ29134 Encodes R
19	306.5	13.5	1859	2	AAx85677 Thermococ
20	305	13.5	2835	2	AAx08141 Hyperther
21	305	13.5	4765	2	AAx08132 Protease
22	305	13.5	4765	2	AAx85670 Pyrococcus
23	305	13.5	4765	2	AAx05921 WO9856926
24	296.5	13.1	3743	9	ADD24901 DNA encod
25	295	13.0	1560	7	ABx37569 Streptomy
26	295	13.0	59816	7	ABx37516 Streptomy
27	295	13.0	59816	7	ABx37515 Streptomy
28	282	12.5	564	2	AAx08134 DNA seque
29	282	12.5	564	2	AAx85676 Thermococ
30	281.5	12.4	898	2	AAx08131 Hyperther
31	281.5	12.4	2166	6	AAx34433 C-termina
32	281.5	12.4	2166	6	ABL40498 DNA const
33	279.5	12.4	546	4	ABL53453 T. yonsei
34	272.5	12.0	4716	9	ABQ80437 CapA codi
35	272.5	12.0	4740	7	ABx28957 Group B S
36	272	12.0	3413	2	AAV72330 F. palust
37	270	11.9	807	2	AAQ11835 Carbonyl
38	269	11.9	810	7	ABx21470 Mature Pa
39	269	11.9	866	2	AAQ30104 Synthetic
40	269	11.9	866	2	AAQ30146 Synthetic
41	269	11.9	866	2	AAQ31697 Subtilisi
42	269	11.9	866	2	AAQ26410 Subtilisi
43	269	11.9	1140	2	AAx37502 Bacillus
44	269	11.9	1140	2	AAx3806 Bacillus
45	269	11.9	1140	2	AAx69598 B. lentus

ALIGNMENTS

RESULT 1

AAQ27516
ID AAQ27516 standard; DNA; 1299 BP.

XX AC AAQ27516;

XX DT 05-FEB-1993 (first entry)

XX DE Alkali-protease Ya enzyme gene.

XX KW Alkali resistance; surface active agent resistance; detergency improver;
XX KW ss.

XX OS Bacillus sp. Y.

XX FH Key Location/Qualifiers

XX FT CDS 1..1299

XX FT /*tag= a

XX PN JP04197182-A.

XX PD 16-JUL-1992.

XX PF 28-NOV-1990; 90JP-00327110.

XX PR 28-NOV-1990; 90JP-00327110.

XX PA (LION) LION CORP.

XX DR WPI; 1992-288440/35.

XX DR P-ESDB; AAR26274.

XX

PT DNA coding alkali-protease Ya enzyme - has good alkali and surfactant

XX resistance and improves detergency.

XX Claim 3; Page 2; 17pp; Japanese.

XX The sequence is that of the alkali-protease Ya enzyme gene which can be

CC used in the recombinant production of Ya enzyme. Ya enzyme is excellent

CC in alkali resistance and surface active agent resistance and improves

CC detergency

XX Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;

Alignment Scores: 2,27e-168 Length: 1299

Pred. No.: 2239.00 Matches: 428

Score: 2239.00 Conservative: 2

Percent Similarity: 99.31% Mismatches: 3

Best Local Similarity: 98.85% Indels: 0

Query Match: 98.94% Gaps: 0

DB: 2

US-09-985-689A-3 (1-433) x AAQ27516 (1-1299)

Qy 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnThrGlyLeuTyr 20

Db 1 AATGAGTGTAGCAAGAGGATAGTAAAGCTGATGTTGTCACAAACAAATACGGATTATAT 60

Qy 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 61 GGCAAGGTCACTAGTTGCGTAGCGGACACAGGCTTAGATACAGGTCGTAAACGATAGT 120

Qy 41 SerMetHisGluAlaPheArgGlyLeuThrAlaLeuTyrAlaLeuGlyValArgThrAsn 60

Db 121 TCATGATGAAGCATTCGCGGGAATAACACAGCTCTTTAGCGTTAGAGAACTAAT 180

Qy 61 AsnAlaAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80

Db 181 AATGCGAGTATCCGAATGGCATGGCCACACATGTAGCAGGTTCTGTACTTGGTAATGT 240

Qy 81 LeuAsnIleGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100

Db 241 TTAATAAAGGAATGGCTCCGCAAGCTTAACCTTTCCTCAATCTAATATGATAGCAGC 300

Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrAsnAla 120

Db 301 GGAGGATTAGTGGCTACCATCGAATTAATACGTTATTTAGTCAAGCTTGGAAATGCT 360

Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaThrAlaAsn 140

Db 361 GGAGCAAGAATTCATACCTCTTGGGGAGCCCCAGTAAATGGAGCGTACACTGCTAAC 420

Qy 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160

Db 421 TCGAGACAAGTGGATGATGTTTCGAAATAATCATATGACGCTACTTTTTCAGCTGT 480

Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIleAsnAlaIleThr 180

Db 481 AATGAAGTCTCTAATTCAGGAACAATTAGTGTCCAGGTACAGCGAATAATGCTATTACG 540

Qy 181 ValGlyValThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200

Db 541 GTCCGCGCACGGAACAACTATCGCCCAAGCTTCGGTTCGATAGCAGTAACCAATCAT 600

Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleIleAspValThr 220

Db 601 ATTCACAAATTTTCATCGAGAGGAGCTACGAGGGATGGACGAATTAAGCCCTGACGTAACA 660

Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240

Db 661 GCTCCTGGAACATTTATTTATCAGCAGCTTCTTCCTAGCTCCAGACTCTTCGTTTGG 720

Qy 241 AlaAsnTyrAsnSerLysTyrAlaThrMetGlyGlyThrSerMetAlaThrProIleVal 260

Db 721 CGCAATTATACAGTAATAACGCGTATATGGCGGTACCTCCATGCGGACACCTATTGTT 780

Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280

Db 781 GCAGGGAAATGTCGCGCAATACGTGAGCATTTTATAAAAAATAGAGGTATTACTCCTAAG 840

Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300

Db 841 CCTTCTTTAATAAAGCTGCACCTTATCGCTGGTGCTACTGATGTTGGTTAGGATATCCT 900

Qy 301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaIleVal 320

Db 901 AGTGGTGACCAAGGCTGGGGGCGTCTCTCTAGATAAATCGTTAAATGTAGCGTATGTC 960

Qy 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340

Db 961 AATGAGCAAACTGCATTAGCCACAGGACAAAAGCAACGCTATTCGTTCCAAGCACAAGCG 1020

Qy 341 GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer 360

Db 1021 GGTAAACCTTTAAANATCTCGTTAGTATGGACAGATGCTCTTGGAASTACACTGCATCT 1080

Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnIleTyrVal 380

Db 1081 TATACACTAGTTAATGATTTAGATCTAGTTATTACTCTCCGAATGACACAAAATATGTA 1140

Qy 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnValGluAsn 400

Db 1141 GGAATGATTTTATGTTATCTTATGATATTAATCTGGATGGATGGTCCGCAACAATGTTGAGAAC 1200

Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaIleVal 420

Db 1201 GTATTTATAACGCTCCGCAATCTGGAACTGATATAATTAATGAGGTTCAAGCGTATAATGTA 1260

Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433

Db 1261 CCATCTGGCCACAGCGCTTCTCCTAGCTATGCTACAT 1299

RESULT 2

AAV82382

ID AAV82382 standard; DNA; 3003 BP.

XX AAV82382;

XX 12-APR-1999 (first entry)

DE Bacillus Jp170 protease gene.

XX Protease; detergent; surfactant; leather processing; debittering;

KW flavour; ss.

XX Bacillus sp.

OS Location/Qualifiers

PH 846..2771

FT CDS

FT sig_peptide

FT mat_peptide

FT 846..944

FT 1470..2768

XX W09856927-A2.

XX 17-DEC-1998.

XX 09-JUN-1998; 98WO-US012005.

XX 12-JUN-1997; 97US-00873479.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX Sloma A, Christianson L;

XX WPI; 1999-080908/07.

DR P-PSDB; AAW89547.
 XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.
 XX Claim 11; Page 52-53; 77pp; English.
 XX This nucleotide sequence encodes a novel protease (see AAW82382) of
 CC *Bacillus* sp. JPI170 (NCIB 12513). The sequence in plasmid p170BAN is
 CC contained in *Bacillus subtilis* LC20 NRRL B-21680. The protease gene was
 CC isolated from chromosomal DNA of JPI170 following preparation of probes
 CC based on protease N-terminal and internal peptides (see AAW89549-50).
 CC screening of chromosomal libraries, isolation of the 3' end of the gene
 CC by inverse PCR (see AAW82410-11) reconstruction of 5' and 3' ends and
 CC PCR amplification (see AAW82412-16). Claimed recombinant host cells can
 CC be used in a method for producing the protease. The protease is used in
 CC laundry and dishwashing detergents, for institutional and industrial
 CC cleaning, and for leather processing, as well as for debittering and
 CC enhancing the degree of hydrolysis of protein hydrolysates, for flavour
 CC development through hydrolysis of proteins, degradation of undesired
 CC peptides and in enzymatic synthesis of peptides. It has enhanced
 CC stability towards oxidation under alkaline conditions, e.g. towards
 CC bleaching agents of the peroxy type. The invention also provides mutant
 CC cells in which the protease activity is diminished. Such cells can be
 CC used for the production of heterologous recombinant proteins
 XX
 SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,83e-152 Length: 3003
 Score: 2040.00 Matches: 385
 Percent Similarity: 94.46% Conservative: 24
 Best Local Similarity: 88.91% Mismatches: 24
 Query Match: 90.15% Indels: 0
 DB: 2 Gaps: 0
 US-09-985-689A-3 (1-433) x AAW82382 (1-3003)
 QY 1 AsnAspValAlaArgGlyLeuValValAlaAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 1470 AATGACGTGGCCGCGGCGATTGTGAAGCAGACGTCGACAAATAAATCTTGGCTTATAT 1529
 QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 1530 GGACAGGACAGATTGTAGCAGTCTGCTGATCTGGGCTTGATACAGGAAGAAATGACAGT 1589
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 1590 TCGATGCATGAACATTCCGCGGTGAAGATTACCGCACTATATGCACTGGCGAGAACGAT 1649
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
 Db 1650 AACGCCAATGATCCAAATGGACATGAACCCCATGTTGCTGATCTGTGTAGGAATGCT 1709
 QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
 Db 1710 ACAATATAAGGATGGCAGCCGACGCCAATCTAGTCTTTCAATCTATTATGGATAGTGGT 1769
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrAsnAla 120
 Db 1770 GGAGGGCTGGGAGGACTACTCTAATCTCAAAACATTATTTCAGTCAAGCATATAGTCT 1829
 QY 121 GlyAlaArgIleHisThrAsnSerThrGlyAlaProValAsnGlyValAlaThrAlaAsn 140
 Db 1830 GGAGCGAGAATTCATACGANNATTCATGGGGGGCTCCAGTAACGGTGCCTATACGACAGAC 1889
 QY 141 SerArgGlnValAspGlyTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
 Db 1890 TCTCGAATATGTTGATGATTATGTGAGAAAAAATGATATGACGATCTTTTTCGCGCGGA 1949
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 1950 AATGAGGGACCGAGGTAGCGGTACAATCAGTGCACCGAGAACAGCAAAATGCGATTACA 2009
 QY 181 ValGlyValAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
 Db 2010 GTTGGGCAACGAAAAACCTAGCTCCAGCTTCGATCTTATCGGATAATATTAAACCAT 2069
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 2070 GTTGCTCAATCTCTTCACAGGTCCTACTAGAGATGGACGATTATAGCCGACGTCATG 2129
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThr 240
 Db 2130 GCACCAAGGTACGTATATTCTCTCTGATAGTATCATCTAGCTCCAGATTCTCTCAATCTGG 2189
 QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
 Db 2190 GCNAACCATGATAGTAATAATGCTCATCGGTGGTACTCTTATGCTACTCCATTGTA 2249
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
 Db 2250 GCAGGTAATGTTGCACAAATAAGGGAGCATTTTGTGAAAAATAGAGGGGTAACTCCTAAG 2309
 QY 281 ProSerIleLysAlaAlaIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
 Db 2310 CCTTCCTTTTAAAGCTCTTAAATGTCAGGTGCTGCGGATGTTGGACTTGCTTTTCCA 2369
 QY 301 SerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
 Db 2370 AATGGTAACCAAGGATGGGGAAGAGTAACGTTAGATAAATCCCTAAATGTCGATTGTG 2429
 QY 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
 Db 2430 AATGAAACGAGCCCTTTATCAACAGTCAAAAGCAACATATTCGTTTACGGCTCAAGCT 2489
 QY 341 GlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrThrAlaSer 360
 Db 2490 GGTAAACCTTAAATAATATCATTGTTTGGTCAGATGCACAGGTAGCAGCGGATCA 2549
 QY 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
 Db 2550 CTAACTTTAGTGAATGATTAGACTTAGTAATCACTGCACCAATGGAATCAATATACGTC 2609
 QY 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGluAsn 400
 Db 2610 GGAATAGCTTTACGACACCGTATGATAACAATTTGGATGGCAGAAACCAACGTGNAAT 2669
 QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
 Db 2670 GTGTTTATCAATGCTCTCTCAAGCGGAACGTATACAGTCGAAGTGCAGGCTTACAATGTA 2729
 QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 2730 CCAGTAAGTCCGCAAACTTTCTTTAGCGATTGTACAT 2768
 RESULT 3
 AAX37277
 ID AAX37277 standard; DNA; 1920 BP.
 XX AC AAX37277;
 XX DT 20-MAR-2003 (revised)
 XX DT 21-JUL-1999 (first entry)
 XX DE *Bacillus* alkaline protease encoding DNA.
 XX KW Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
 XX KW washing composition; oxidising agent; ss.
 XX OS *Bacillus* sp.
 XX PN WO9918218-A1.
 XX XX 15-APR-1999.
 XX PD

```

07-OCT-1998;   98WO-JP004528.
XX
XX
PR OCT-1997;   97JP-00274570.
XX
PA (KAOS ) KAO CORP.
PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
PI Shikata S, Nomura M;
DR WPI; 1999-287736/27.
XX P-PSDB; AAV17087, AAV17089.
XX Alkali protease from Bacillus used in washing powders.
PT
XX Disclosure; Page 53-58; 7lpp; Japanese.
XX
XX The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can
CC be used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease encoding
CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
XX
XX Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      4.35e-149      Length:      1920
Score:          1998.50        Matches:     382
Percent Similarity: 94.01%    Conservative: 26
Best Local Similarity: 88.02%  Mismatches:  25
Query Match:     88.31%       Indels:      1
DB:              2           Gaps:         1

US-09-985-689A-3 (1-433) x AAX37277 (1-1920)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 616 AATGATGTGGCAGAGGTATTGTCAAAGCGGATGTGCACAGAGCACTACGGTTTGAT 675
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAGCCAGCATGTTCGGAGTTGCCGACTCGGATCGGATACAGAGAAGAACGACAGT 735
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGTCATGAAGCCCTTCGCGCGGTAAATAACAGCACATATATGCTCGGTGCGAGCAAT 795
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValalaglySerValLeuGlyAsn--- 79
Db 796 AATCGCAATGTATACGAACGGTCA'GGTACCCTATGGCAGGTCGGTATTAGGAATGCG 855
Qy 80 AlaLeuAsnLysGlyMetAlaproGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCACGCAATAAGGAATGGCACCTCAGCGAATCTGGTTTTTCAATCCATCATGTGATAGC 915
Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 916 AGTGGTGGGCTTGAGGGTCTGCCCTTCCAATTCGAAACCTTTATTCAGCCAAAGCATTCAGT 975
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTirpGlyAlaProValAsnGlyValaTyThrAla 139
Db 976 CGAGGTGCGAGAAATTCATAAACTCTGGGGGCGACGGGTGATGGGCTTACACGACA 1035
Qy 140 AsnSerArgGlnValAspGlyTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159

```

XX WO9918218-A1.
PN 15-APR-1999.
PD 07-OCT-1998; 98WO-JP004528.
XX 07-OCT-1997; 97JP-00274570.
PR (KAOS) KAO CORP.
PA
XX
XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
PI Shikata S, Nomura M;
XX WPI; 1999-287736/27.
DR P-PSDB; AA17091.
XX
XX Alkali protease from Bacillus used in washing powders.
PT
XX
XX Disclosure; Page 63-68; 71pp; Japanese.
XX
XX The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease encoding
CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
XX
SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7.55e-149 Length: 1923
Score: 1995.50 Matches: 380
Percent Similarity: 94.01% Conservative: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.18% Indels: 1
DB: Gaps: 1

US-09-985-689A-3 (1-433) x AAX37279 (1-1923)

QY 1 AsnAspValAlaArgGlyIleVallyAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTTGCGGTGGAATTTGCAAGCGGATGGCTCAGCAGCAGCTACGGTTGTAT 678
QY 21 GlyGlnGlyGlnValAlaAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAGGACAGATCGTACGGTTGCCGATACAGGCTTGTATACAGGTGCGCAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyValGthrAsn 60
Db 739 TCGATGATGAGACCTTCGCGGGAAATTAATCTGCAATTATATGCTTGGACGCGAGAT 798
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATGATACGAAATGGTTCATGTGTACGATGTGGCTGGCTCGGCTATTAGGAACCGC 858
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCAATCATCATGATAGC 918
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 919 GGTGGGGGACTTGGAGGACTACCTTCGAATCTGCAAACTTATTACGCCAAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139

Db 979 GCTGGTGGCAGATTCATCAAACTCCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCAGAAATGTGATGATGATGCGCAAAATGATGATGATGATGATGATGATGATGAT 1098
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGGATGAGGACCGAAGCGGGAACCATCAGTGCACAGGACACAGCAATGAATGAATA 1158
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACAGTGGAGCTACGGAAACCTCCGCCCAAGCTTTGGTCTTATCGGACCAATATCAAC 1218
QY 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGCACAGTCTCTTCACGTGCACCGCAAGCATGGACGATCAACCGGATGTC 1278
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGCCACCGGAAACGTTTCTATCATCAGCAAGATCTTCTCTGACCGGATTCCTCTTC 1338
QY 240 TrpAlaAsnTyrAsnSerIleTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGAAGCACTGACAGTAATATGATCATCATGGTGGAACTGCTCATGGCTACACGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1399 GTTGTGGAACCGTGGCACAGCTTCGTGAGCATTTGTGAAAAACAGAGCATCACACCA 1458
QY 280 LysProSerIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCTTCTCTATTAAGCGGACATGATTCGCGGTGACGTACATCGGCTTGCTGCTAC 1518
QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGAAGGTAAACCAAGGATGGGACGAGTGACATTCGATAAATCCCTGAAGCTTGCTAT 1578
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1579 GTGAACGAGTCCAGTTCTCTATCCACGCAAAAGCGACGCTACTCGTTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
Db 1639 GCGGCGAGCGCTTGAATAATCTCCCTGGTATGCTGATGCCCTCGGAGCACACTGCT 1698
QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCCGTAAACGCTTGTCAATGATCTGGACCTTGTCAATACCGCTCCAAATGCGCACAGTAT 1758
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGlu 399
Db 1759 GTAGAAATGACTTACTTCGCAATCATATGATTAATGGATGGCGGCAATACGTAGAA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGATTTAATTAATGACCAAAAGCGGACGCTATACAAATGAAGTACAGGCTTAAAC 1878
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGTTGGACACAGAACTTCTCGTTGCAATTTGTGAAT 1920
RESULT 5
AAX37278
ID AAX37278 standard; DNA; 1923 BP.
XX AAX37278;
AC AAX37278;
XX
XX 20-MAR-2003 (revised)
DT 21-JUL-1999 (first entry)
XX
DE Bacillus alkaline protease encoding DNA.

XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX Bacillus sp.
 XX WO9918218-A1.
 XX 15-APR-1999.
 PD 07-OCT-1998; 98WO-JP004528.
 FF 07-OCT-1997; 97JP-00274570.
 XX (KAOS) KAO CORP.
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17088, AAY17090.
 XX PT Alkali protease from Bacillus used in washing powders.
 XX Disclosure; Page 58-63; 71pp; Japanese.
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX SQ Sequence 1923 BF; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 9,06e-149 Length: 1923
 Score: 1994.50 Matches: 380
 Percent Similarity: 94.01% Conservative: 28
 Best Local Similarity: 87.56% Mismatches: 25
 Query Match: 88.14% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689a-3 (1-433) x AAX37278 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValIleAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 DB 619 AATGATGTTCGCGTGGAAATGTCAAAGCGGATGTGCTCAGACGACTACGGGTGTAT 678

QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 679 GGACNAGACAGATCGTAGCGGTTCGCATACAGCGCTTGATACAGTCCGATGACAGT 738

QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 739 TCGATGCATGAAGCCCTCCCGGGGAAATTAAGTGCATATATATGATTCGGACGGACGAAT 798

QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 DB 799 AATGCCAATGATACGAATGGTATCGATCGCATGTGGCTGCTCCGATTTAGGAAACGGC 858

QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 DB 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAACTATCATCGATGATGC 918

QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
 DB 919 GGTGGGGGACTTGGAGGACTTACCTTCGAATCTCAAAACCTTATTACGCAAGCATACAGT 978

QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139
 DB 979 GCTGTGCCAGAAATTCATAAACTCTCTGGGGAGCAGCAGTGAATGGGCTTACACACA 1038

QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
 DB 1039 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCCTTTTCGCTGCC 1098

QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 DB 1099 GGGAAATGAAGACCGAAGCGGGAACCATCAGTGCACCGACGACGATCAAAATGCAATA 1158

QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
 DB 1159 ACAGTCGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACAAATATCAAC 1218

QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
 DB 1219 CATGTGCACAGTTCTCTTCACGTGGACCGCAAGATGCGGATCAAAACCGGATGTC 1278

QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 DB 1279 ATGCACCGGAACGTTTCATCATATCAGCAAGATCTTCTTGCACCGGATTCCTCTTC 1338

QY 240 TrpAlaSerTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
 DB 1339 TGGCGCAACCATGACAGTAAATATGCATATCATGTGGTGAACGTCCTCATCCGATC 1398

QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 DB 1399 GTTGCTGGAACGTTGGACGACGCTTCGTGAGCATTTCTGGAACACAGGACATCACACA 1458

QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaThrAspValGlyLeuGlyTyr 299
 DB 1459 AAGCTTCTCTATTAAAGCGGCATGATTGCCGTGCAGCTGACATCGGCTTGGGTAC 1518

QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 DB 1519 CGGAACGGTAACCAAGGATGGGACGAGTGACATTTGGATTAATCCCTGAACGTTGCCTAT 1578

QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
 DB 1579 GTGACGAGTCCAGTCTCTTATCCACGACCAAAAGCGACGCTACTCGTTTACTGCTACT 1638

QY 340 AlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrThrAla 359
 DB 1639 GCCGCAAGCCCTTGAAATCTCCCTGTGTGGTCTGATGCCCTTCGAGACCAACTGCT 1698

QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 DB 1699 TCCGTAACGCTTGTCATGATCTGACCTTGTTCATTACCGCTCCAAATGGCACACAGTAT 1758

QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 399
 DB 1759 GTAGGAATGACTTTTACTTCGCATACATATACCTGGATGGCGGCAATAAGTAGAA 1818

QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValIleGlnAlaTyrAsn 419
 DB 1819 AATGATTTTATTAATGCCCAAGCGGACGATATACAATGAGTACAGCTTATTAAC 1878

QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 DB 1879 GTACCGTTGGACCAACAGACCTTCTCGTTGGCAATTTGTAAT 1920

RESULT 6
 AAT85667
 ID AAT85667 standard; DNA; 1977 BP.
 XX
 AC AAT85667;

XX 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX Thermococcus protease coding sequence.
 XX Protease; research reagent; thermal stability; thermococcus celer; ss.
 KW Thermococcus celer; DSM-2476.
 OS WO9721823-A1.
 PN 19-JUN-1997.
 XX 07-NOV-1996; 95WO-JP003253.
 XX 12-DEC-1995; 95JP-00323285.
 XX (TAKI) TAKARA SHUZO CO LTD.
 PA Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 PI Tsunawasa S, Kato I;
 XX WPI; 1997-332794/30.
 DR P-PSDB; AAM24121.
 XX
 PT Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX Claim 3; Page 86-87; 159pp; Japanese.
 XX This sequence represents the coding sequence for the protease from
 CC Thermococcus celer DSM-2476. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: Length: 1,098-25 1977
 Score: 443.00 Matches: 137
 Percent Similarity: 46.03% Conservative: 66
 Best Local Similarity: 31.07% Mismatches: 170
 Query Match: 19.58% Indels: 68
 DB: 2 Gaps: 16
 US-09-985-689A-3 (1-433) x AAT85667 (1-1977)
 QY 8 ValLysAlaAspValAlaGlnAsnSerGlyLeuTyrGlyGlnValValAla 27
 DB 433 ATAGGGGCGGATACCGTCTGGAACTCCCTCGGCTACGACGGAAGCGGTGGTGGTGC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
 DB 493 ATCGTCGATACGGGTATAGACCGAAC-----CACCCCGATCTGAAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAspProAsn 66
 DB 535 GGCAAGGTATAGGTGGTACACGCCGCTCAACGCGCAGGTCCGCCGAGTACGCCAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
 DB 595 GGACACGGAAACCAAGTTCGGGGTATCGTTGCCGGAACGGGAGCGGTAACTCCAGTAC 654
 QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
 DB 655 ATAGGCGTCCGCCCGCGGCGAAGTCTCGTGGCGCTCAAGGTTCTCGTGGCGGAGCGGTTCG 714
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120

DB 715 GGAAGCGTCTCCACCATCATCGCGGGTGTGACTGGTGGTCTCCAGAACAGACAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 DB 775 GGGATAAGGGTATCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGGAAACCGACTCC 834
 QY 141 SerArgGlnValAspGluTyrValArgAsnAsnMetThrValLeuPheAlaAlaGly 160
 DB 835 CTCAGTCAGCGCGTCAACACCGCTGGGACGCGGTATAGTAGTCTGCGTCCGCCCGGC 894
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 DB 895 AACAGCGGGCGGAACACCTACACCGTCCGCTCACCGCGCGCGAGGAGTCTATAACC 954
 QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
 DB 955 GTCCGTGCA-----GTTGACAGCAGCAGCAAC 981
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 DB 982 ATGCCAGCTTCTCCAGCAGGGACCGACCGCGGACGGAAGGCTCAAGCGGAGTCTGTC 1041
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 DB 1042 GCCCGCGGGTTCACATCATAGCCCGCGCGCAGC-----GGAACAGCATGGGC 1092
 QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
 DB 1093 ACCCGCATAAACGACTACTACACCAAGGCTCTGGAACCCAGCATGCCACCCGCGACGTT 1152
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
 DB 1153 TCGGGCGGTGGCGCGTCTATCTCTCCAG-----GCCACCCGAGCTGGAGCCCGGAC 1203
 QY 280 -----LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
 DB 1204 AAGGTGAAGACCGCCCTCATCGAGACCGCGACATGTCGCCCCCAAGAGATAGCGGAC 1263
 QY 297 LeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsn 316
 DB 1264 ATGCCCTAC-----GGTGGGTAGGGTGAACGCTCTACAAGGCC----- 1302
 QY 317 ValAlaTyrValAsnGluAlaThrAlaLeuThrThrGlyGln-----Lys 331
 DB 1303 ATCAAGTACGACGACTACGCCAAGCTACCTTCACCGGCTCCGTCGCCGCAAGGGAAGC 1362
 QY 332 AlaThrTyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTrpThr 351
 DB 1363 GCCACCCACACCTTCGACGTCAGCGCGCCACCTTCGTGACCGCCACCTCTACTGGGAC 1422
 QY 352 AspAlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuValIle 371
 DB 1423 -----ACGGGCTCGAGCGACATCCACTCTACCTC 1452
 QY 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
 DB 1453 TACGACCCCAAC-----GGGAACGAGGTTCAGCTACTCTCTACACCGCGCTAC 1497
 QY 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
 DB 1498 TAC-----GGTTTCGAGAAGGTTCGGCTACTACAACCCGACCGCGGAGACTGG 1545
 QY 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 DB 1546 ACGGTCAAGTCTGTAGTAC-----AAGGGCGCGGGAACCTACCAGGTCCGACGTC 1596
 QY 432 Val 432
 DB 1597 GTC 1599
 RESULT 7
 AAX05926
 ID AAX05926 standard; DNA; 1977 BP.

XX AAX05926;
 AC
 XX
 DT 06-MAY-1999 (first entry)
 XX
 DE WO9856926 Seq ID 11.
 XX
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 XX
 OS Thermococcus celer.
 XX
 PN WO9856926-A1.
 XX
 PD 17-DEC-1998.
 XX
 PF 04-JUN-1998; 98WO-JP002465.
 XX
 PR 10-JUN-1997; 97JP-00151969.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX
 PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 XX
 DR WPI; 1999-080907/07.
 XX
 DR P-PSDB; AAW94840.
 XX
 PT Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.
 XX
 PS Disclosure; Page 53-54; 82pp; Japanese.
 XX
 CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula Sig-Ala-Gly-Asn-Pro, where Sig is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis
 XX
 SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,09e-25 Length: 1977
 Score: 443.00 Matches: 137
 Percent Similarity: 46.03% Conservative: 66
 Best Local Similarity: 31.07% Mismatches: 170
 Query Match: 19.58% Indels: 68
 DB: 2 Gaps: 16
 US-09-985-689A-3 (1-433) x AAX05926 (1-1977)
 QY 8 VallysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyGlnValValAla 27
 DB 433 ATAGGGGCCGATACCGTCTCGAATCTCCCTCGGTACGAGGAGCGGTGTGGTGTGCC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethHisGluAlaPheArg 47
 DB 493 ATCGTCGATACGGGTATAGACGGGAAC-----CACCCCGATCTCGAAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
 DB 535 GGCAGGTCTAGGCTCGTACGACGGCGTCAACGGCAGGTCTGACCCCTACGATGACCCAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
 DB 595 GGACACGGAAACCCACGTTGGGGTATCGTTGCCGGAAACCGGACCGGTTAACTCCCAAGTAC 654

QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
 DB 655 ATAGGGCTCGCCCCCGCGCGAAGCTCGCGCGTCAAGGTCTTCGGTCCCGACGGTTCG 714
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
 DB 715 GGAAGCGTCTCCACCATCATCGGGGTGTGATCGGTCTGTCAGAACAGGACAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 DB 775 GGGATAAGGGTCAATCAACCTCTCCCTCGGTCTCTCCAGAGTCCCGACGGAACCGACTCC 834
 QY 141 SerArgGlnValAspGluTyrValA:GAsnAsnAspMetThrValLeuPheAlaAlaGly 160
 DB 835 CTCAGTCAGCGCGCTCAACAACGCTGGAGCGCGGTATAGTAGTCTCGCTCGCGCGCGC 894
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 DB 895 AACAGCGGGCGGAACACTACACCGTGGTCCCGCGCGCGAGGAGTCAAGTCAATAACC 954
 QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
 DB 955 GTCGGTGCA-----GTTGACGACGACGACCAAC 981
 QY 201 IleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 DB 982 ATGCCAGCTTCTCCAGAGGGACCGACCGCGAGCGAAGGCTCAAGCGGAGTCTCGTC 1041
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 DB 1042 GCGCGCGGTGACATCATATGCGCGCGCGCGCAGC-----GGAACGACATGGGC 1092
 QY 241 AlaAsnTyrAsnSerIstYrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
 DB 1093 ACCCGGATAAAGCACTACTACCAAGCGCTCTGGAACGACGATGCGCACCGCGCAGTT 1152
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
 DB 1153 TCGGGCGTGGCGGCTCATCTCTCCAG-----GCCACCCGAGTGGACCGCGGAC 1203
 QY 280 -----LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
 DB 1204 AAGGTGAAGACCGCCCTCATGAGCGCGCATAGTCCGCCCCCAAGGAGATAGCGGAC 1263
 QY 297 LeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsn 316
 DB 1264 ATCGCCTAC-----GGTGGGTGGGTGAACGCTCTACAGGCGC----- 1302
 QY 317 ValAlaTyrValAsnGluAlaThrAlaLeuThrThrGlyGln-----Lys 331
 DB 1303 ATCAAGTACGACGACTAGCGCAAGCTCACCTTCACCGGCTCCGTCGCGCAAGGGAAGC 1362
 QY 332 AlaThrTyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTrpThr 351
 DB 1363 GCCACCCACACCTTCGACGTCAGCGCGCCACCTTCGTGACCGCCACCTCTCTACTGGAC 1422
 QY 352 AspAlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle 371
 DB 1423 -----ACGGGCTCGAGCGACATCGACCTTACCTC 1452
 QY 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
 DB 1453 TACGACCCCAAC-----GGGAACGAGGTGTGACTACTCTCTACACCGCTAC 1497
 QY 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
 DB 1498 TAC-----GCGTTCGAGAGGTGGCTACTACACCGCGCGCGGACCTGG 1545
 QY 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 DB 1546 ACGGTCAAGGTCTGTCAGTAC-----AAGGCGCGCGGCAACTACCGAGGTCTGACGTC 1596


```

QY 432 Val 432
Db 1597 GTC 1599

RESULT 8
AAK05920
ID AAX05920 standard; DNA; 1236 BP.
AC AAX05920;
XX
DT 06-MAY-1999 (first entry)
XX
DE Hyperthermostable protease fragment encoding DNA.
KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
XX
OS Pyrococcus furiosus.
XX
PN WO9856926-A1.
XX
PD 17-DEC-1998.
XX
PF 04-JUN-1998; 98WO-JP002465.
XX
PR 10-JUN-1997; 97JP-00151969.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
FI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
DR WPI; 1999-080907/07.
XX
DR P-PSDB; AAW94836.
XX
PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and
PT gene encoding it, for large scale production of the protease for
XX industrial use.
XX
PS Claim 6; Page 37-38; 82pp; Japanese.
XX
CC The invention relates to a hyperthermostable protease derived from a
CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
CC 95 deg.C. The invention also provides gene sequences encoding a
CC polypeptide of formula SIG-Ala-Gly-Ans-PRO, where SIG is a signal
CC peptide from subtilisin, and PRO is the above protease. Host cells
CC (especially Bacillus strains) transformed with vectors comprising the
CC genes are used for the recombinant production of the protease. The
CC hyperthermostable protease which can be prepared in quantity suitable for
CC industrial use, can be used as an additive for drugs, washing agents and
CC foodstuffs and for chemical synthesis
XX
SQ Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,6e-24 Length: 1236
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.67% Indels: 79
DB: 2 Gaps: 17

US-09-985-689A-3 (1-433) x AAX05920 (1-1236)
QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTCGGATATGATGTTCTGGAATCAATAGGAATAATGACACTGGAATTGAC--- 114
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyValLeuThrAlaLeuTyrAla 55
Db 115 -----GCTTCTCATCCAGATCTCCAGGAAGAAGTA----- 144

```

Db 1096 GCTACTATGATTCGAAAGGTTGGTTATTACAACCCAACTGATGGAACATGGCAAT 1155
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1156 AAGTTGTGAAGCTAC-----ACGGGAAGTGCAAACTATCAAGTAGATGTGTA 1203

RESULT 9

AAT85668 standard; DNA; 1566 BP.

XX AC AAT85668;
 XX 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX Pyrococcus furiosus protease coding sequence.
 DE
 XX Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
 KW
 XX Pyrococcus furiosus; DSM-3638.
 OS
 XX Key Location/Qualifiers
 FH CDS
 FT 1.1566
 FT /tag= a
 FT /transl_except= (pos: 1282..1284, aa: Xaa)
 FT /note= "Xaa= Gly, Val"

WO9721823-A1.

19-JUN-1997.

07-NOV-1996; 96WO-JP003253.

12-DEC-1995; 95JP-00323285.

(TAKI) TAKARA SHUZO CO LTD.

Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;

Teunasaawa S, Kato I;

WPI; 1997-332794/30.

P-PSDB; AAW24122.

XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX Claim 7; Page 90-91; 159pp; Japanese.

XX This sequence represents the coding sequence for the protease from
 CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)

SQ Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 3,47e-24 Length: 1566
 Score: 422.50 Matches: 138
 Percent Similarity: 43.96% Conservative: 55
 Best Local Similarity: 31.44% Mismatches: 167
 Query Match: 18.67% Indels: 79
 DB: 2 Gaps: 17

US-09-985-689A-3 (1-433) x AAT85668 (1-1566)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35

Db 58 AACTTGGATATGATGTTCTGGATCACAATAGGAATATTGACACTGGATTTGAC--- 114

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 115 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 144
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 145 ATTGGGTGGGTAGATTGTCATGTTAGGAGTTATCCATACGATGACCATGGACATGGA 204
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAAGGA 264
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 265 ATGGCTCCAGGAGCTAAGCTGGCGGGAATTAAAGTTCTAGTCCCGATGGTTCTGGAGC 324
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 325 ATATCTACTATAATTAAGGAGTTGAGTGGCGCGTTGATTAACAAGATAGTAGCGAATT 384
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 385 AAGTCAATTAATCTTCTCTGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 444
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 445 CAGGCTGTTAATGAGCGCTGGGATCTGGATTAGTTGTTGTTGCGCTGGAAACAGT 504
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 505 GGACCTAACAAAGTATCAATCGGTTCTCCAGAGCTGCAAGCAAAAGTTATTACATTGGA 564
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
 Db 565 GCCGTTGACAAGTAT-----GATGTTATAACA 591
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 592 AGCTTCTCAAGCAGAGGGCAACTGCAGCGGAGGCTTAAGCCTGAGGTGTTGTCTCCA 651
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 652 GGAACTGGATAATTGCTGCCAGAGCAAGT-----GGAACTAGCATGGGTCAACCA 702
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 703 ATTAATGACTATTACACAGCAGCTCTCTGGAGACATCAATGGCAACTCTCTCAGTAGTGT 762
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 763 ATTGCAGCCCTCTGCTCCAA-----GCACACCCGAGCTGGATCCAGACAAAGTA 813
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly----LeuGly 298
 Db 814 AAAACAGCCCTCATAGAAACTGCTGATATCGTAAAGCCAGATGAATAGCGATATATGCC 873
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 874 TAC-----GGTCAGGTAGGGTTAATGCATACAGAGGTATAAAC----- 912
 QY 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
 Db 913 TACGATAACTATGCAAAAGCTAGTGTTCACCTGGATATGTCGCAACAAAGGCGAGCAAACT 972
 QY 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
 Db 973 CACCAGTTCGTTATTAGCGAGCTTCGTTGTAACCTCCACATTTACTGGCAATGCC 1032
 QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1033 AAT-----AGCGACCTTGATCTTTACCTCTACCAT 1062
 QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393

```

Db      1063 CCCATGGAAACAGGTT-----GACTACTCTTAC-----ACC 1095
Qy      394 GlyArgAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
Db      1096 GCCTACTATGATTGCAAAAGTGTTATTACAAACCACTGATGGAACATGACAAATT 1155
Qy      414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db      1156 AAGTTTGAAGCTAC-----AGCGAAGTGCAAACTATCAAGTAGATGTGGTA 1203

RESULT 10
AAT85695
ID      AAT85695 standard; DNA; 1962 BP.
XX
AC      AAT85695;
XX
XX      17-OCT-2003 (revised)
DT      20-APR-1998 (first entry)
XX
XX      Pyrococcus furiosus PFUS protease coding sequence.
XX
KW      Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX
OS      Pyrococcus furiosus; DSM-3638.
XX
FN      WO9721823-A1.
XX
PD      19-JUN-1997.
XX
XX      07-NOV-1996; 96WO-JP003253.
XX
XX      12-DEC-1995; 95JP-00323285.
XX
XX      (TAKI ) TAKARA SHUZO CO LTD.
XX
XX      Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
PI      Tsunawasa S, Kato I;
XX
XX      WPI; 1997-332794/30.
DR      P-PSDB; AAM24129.
XX
XX      Protease(s) and genes encoding them obtained from Thermococcus and
PT      Pyrococcus strains - have extremely high thermal stability and are useful
PT      industrially and as research reagents.
XX
XX      Disclosure; Page 123-125; 159pp; Japanese.
XX
XX      This sequence represents the coding sequence for the protease from
CC      Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
CC      invention. The proteases of the invention have extremely high thermal
CC      stability. The proteases can be used as research reagents, and
CC      industrially in the food, drug and chemical industries. (Updated on 17-
CC      OCT-2003 to standardise OS field)
XX
SQ      Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      4,569-24      Length:      1962
Score:          422.50      Matches:      138
Percent Similarity: 43.96%      Conservative: 55
Best Local Similarity: 31.44%      Mismatches: 167
Query Match:     18.67%      Indels:      79
DB:             2           Gaps:        17

US-09-985-689A-3 (1-433) x AAT85695 (1-1962)

Qy      16 AsnTyrGlyLeuTyrGlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThr 35
Db      454 AACTTGGGATATGATGGTTCTCGAATCAATAGGAATAATTGACACTGGAATTGAC--- 510
Qy      36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55

```

```

511 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 540
Qy      56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db      541 ATTGGTGGGTAGATTTTGTCAATGGTAGAGTATCCATACGATCACCATGCATGGA 600
Qy      70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db      601 ACTCATGTAGCTTCAATAGCAGCTGTACTGGAGCAGCAAGTAATGGCAAGTACAGGA 660
Qy      85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db      661 ATGGCTCCAGGAGCTAAGCTGCGGGAATTAAAGTTCTAGGTGCGGATGGTCTTCTGAAGC 720
Qy      103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAla 122
Db      721 ATATCTACTATATAATTAAAGGAGTTGAGTGGCGGTGATTAACAAGATAAAGTACGGAATT 780
Qy      123 ArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db      781 AAGGTCAATTAATCTTTCTCTGTTCAAGCCAGAGCTCAGATGCTGACTGACCTCTAAGT 840
Qy      143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db      841 CAGGCTGTTAATGCACGCGTGGATGCTGATTAGTTGTTGTTGCTGCCCTGGAACAGT 900
Qy      163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db      901 GGACCTAAACAAGTATACATCGTTCTCCAGCAGCTGCAAGCAAGATTATTACAGTTGA 960
Qy      183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db      961 GCGTTTGACAAAGTAT-----GATGTTATAACA 987
Qy      203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db      988 AGCTTCTCAAGCAGAGGCGCACTGCAGACGCGAGCTTAAGCTGAGTTTGTCTCCA 1047
Qy      223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db      1048 GGAAACTGGATAATTGCTGCCAGAGCAAGT-----GGAACTAGCATGGTGTCAACCA 1098
Qy      243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db      1099 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCACTCCTCAGTAGTGTGT 1158
Qy      263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db      1159 ATTGCAGCCCTCTTGTCTCAA-----GCACACCCGAGCTGGACTCCACACAAAGTA 1209
Qy      280 LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db      1210 AAAACAGCCCTCATAGAAACTGCTGATATCGTAAAGCCAGATGAATAGCCGATATAGCC 1269
Qy      299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db      1270 TAC-----GGTGCAGTAGGTTTAATGCATACAGGCTATAAAC----- 1308
Qy      319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
Db      1309 TACGATAACTATGCAAAAGCTAGTGTTCCTGATGATATGTTGCCAACAAGGAGCCAAACT 1368
Qy      334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTyrThrAspAla 353
Db      1369 CACAGTTGTTATTAGCGGAGCTTCGTTCTGTAACGCCACATTATCTGGGCAATGCC 1428
Qy      354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db      1429 AAT-----AGCGACCTTGATCTTTTACCTCTACAGT 1458
Qy      374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db      1459 CCCAATGGAAACAGGTT-----GACTACTCTTAC-----ACC 1491

```

QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
 Db 1492 GCTACTATGATTCGAAAGGTGGTTATTACACCCCACTGATGGACATGGACATT 1551
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1552 AAGTTGTAAGCTAC-----AGCGAAGTGCAAACTATCAAGTAGATGTGTA 1599

RESULT 11

AA05929
 ID AAX05929 standard; DNA; 1962 BP.

XX AC AAX05929;

XX DT 06-MAY-1999 (first entry)

XX DE Hyperthermostable protease encoding DNA.

XX KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 XX additional; drug; washing agent; foodstuff; chemical synthesis; ds.

XX OS Pyrococcus furiosus.

XX PN WO9856926-A1.

XX PD 17-DEC-1998.

XX PF 04-JUN-1998; 98WO-JP002465.

XX PR 10-JUN-1997; 97JP-00151969.

XX XX (TAKI) TAKARA SHUZO CO LTD.

XX PA Takakura H, Morishita M, Shimojo T, Asada K, Kato I;

XX PI WPI; 1999-080907/07.

XX DR P-PSDB; AAW94841.

XX PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 XX gene encoding it, for large scale production of the protease for
 XX industrial use.

XX PS Disclosure; Page 59-60; 82pp; Japanese.

XX CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

XX SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4,56e-24	Length:	1962
Score:	422.50	Matches:	138
Percent Similarity:	43.96%	Conservative:	55
Best Local Similarity:	31.44%	Mismatches:	167
Query Watch:	18.67%	Indels:	79
DB:	2	Gaps:	17

US-09-985-689A-3 (1-433) x AAX05929 (1-1962)

QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 454 AACTTGGGATGATGTTCTTGGATCACAATAGGAATAATTGACATGGATTGAC--- 510

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 511 -----GCTTCTCATCCAGATCTCCAAAGGAAAGTA----- 540
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 541 ATTGGGTGGGTAGATTGTCATGTTAGTGGAGAGTTATCCATAGCATGACCATGACATGA 600
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 601 ACTCATGTAGCTTCAATAGCAGCTGCTACTGGAGCAGCAAGTAATGGCAAGTACAAGCGA 660
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 661 ATGCTCCAGAGCTAAGCTGGCGGAATTAAGTTCTTAGTCCGATGTTCTTGAAGC 720
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 721 ATATCTACTATATAATTAAGGGAGTTGAGTGGCGCTTGATACAAAGATAAGTACGGAAT 780
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 781 AAGTCTAATTAATCTTCTCTGTTTCAAGCCAGAGCTCAGATGTTACTGACGCTCTAAGT 840
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 841 CAGGCTGTTAATGCGAGCTGGATGCTGGATTAGTTGTTGCTGCTGCGCTGGAACAGT 900
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 901 GGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGA 960
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
 Db 961 GCGCTTGCAAGTAT-----GATGTTATAACA 987
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 988 AGCTTCTCAAGCAGAGCGCCAACTGCAGACGGCAGGCTTAAGCCTGAGTTGTTGCTCCA 1047
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 1048 GGAACCTGGATAATTGCTGCCAGAGCAAGT-----GGAACCTAGCATGGTCAACCA 1098
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1099 ATTAATGACTATTACAGCAGCTCTGCGACATCAATGGCAACTCTCCTCAGTAGTGT 1158
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 1159 ATTGCGCCCTCTTGTCTCCAA-----GCACCCCGAGCTGGACTCCAGACAAAGTA 1209
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
 Db 1210 AABACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATAGCCGATATAGCC 1269
 QY 299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 1270 TAC-----GGTGCAGGTAGGTTAATGCTATACAGGCTATAAAC----- 1308
 QY 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
 Db 1309 TACGATAACTATGCAAGCTAGTGTCTCCTGATATGTTGCCAACAAAGGCGAGCAAACT 1368
 QY 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThrAspAla 353
 Db 1369 CACCAGTTTCGTTATTAGCGGAGCTTCGTTCTGTAACCTGCCACATTATATGGCAATGCC 1428
 QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1429 AAT-----AGCAGCTTGATCTCTTACCTCTACGAT 1458

```
QY 374 ProAnglyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db 1459 CCCAATGGAAACAGGTT-----GACTACTCTTAC-----ACC 1491
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
Db 1492 CCCTACTATGATTGCAAAAGGTTGGTTATTATACACCCCACTGATGGAATGGACAATT 1551
QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGGTTGTAAGCTAC-----AGCGAAGTGCAAACTATCAAGTAGATGTGGTA 1599

RESULT 12
AA85669
ID AA85669 standard; DNA; 1977 BP.
XX
AC AA85669;
DT 20-APR-1998 (first entry)
XX
DE Protease coding sequence.
XX
KW Protease; research reagent; thermal stability; ss.
XX
OS Synthetic.
XX
PN WO9721823-A1.
XX
PD 19-JUN-1997.
XX
PF 07-NOV-1996; 96WO-JP003253.
XX
PR 12-DEC-1995; 95JP-00323285.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
PI Tsunawasa S, Kato I;
XX
DR WPI; 1997-332794/30.
XX
P-PSDB; AAW24123.
XX
PT Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.
XX
PS Claim 11; Page 95-97; 159pp; Japanese.
XX
CC This sequence represents the coding sequence for a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries
XX
SQ Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,138-22 Length: 1977
Score: 401.50 Matches: 130
Percent Similarity: 43.96% Conservative: 63
Best Local Similarity: 29.61% Mismatches: 167
Query Match: 17.74% Indels: 79
DB: 2 Gaps: 17

US-09-985-689A-3 (1-433) x AA85669 (1-1977)
QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTTGGGATATGATGTTCTCGAATCAATPAGGAATAATTGACACTGGAATTGAC--- 510
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----GCTTCTCATCCAGATCTCCAAAGGAAAGTA----- 540
```

```
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATTGGTGCGGTAGATTGTTGCAATGGTAGAGTTATCCATAGATGACCATGGACATGGA 600
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 601 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAGGA 660
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 661 ATGGCTCCAGGAGCTAAGCTGCGGGAATTAAAGGTTCTAGGTGCGCATGGTTCTCGAAGC 720
QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 721 ATATCTACTATAATTAAAGGAGTTGAGTGGGCGGTTGATACAAAGATAAGTACGGAATT 780
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 781 AAGGTCATTAAATCTTCTCTCTCAAGCCAGAGCTCCGACGGAACGAGCTACCTCAGT 840
QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaGlyAsnGlu 162
Db 841 CAGGCGCTCAACACGCGTGGGACCGCGGTATAGTAGTCTGCGTGGCGCGCGCAACAGC 900
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 901 GGGCCGAACACCTACACGCTCGCTCACCCTCCGCGCGGAGCAAGGTCAATACCGTCTG 960
QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 961 GCA-----GTTGACAGCAAGCAACATCGCC 987
QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 988 AGCTTCTCCAGCAGGGGACCGCGGACGAGGCTCAAGCCGGAAGTCTGCGCCGCC 1047
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1048 GCGTTTGACATCATATCCCGCGCGCAGC-----GGAACCCAGCATGGCCACCGCC 1098
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrAlaThrAlaGly 262
Db 1099 ATAAACGACTACTACACGAGCGCTCTGGAACACGATGGCCACCGCCGACGTTTGGGC 1158
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 1159 GTTGGCGCGCTCATCTCCAG-----GCCACCGAGCTGGACCCCGCAAGGTG 1209
QY 280 LysProSerIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 1210 AAGACCGCTCATCGAGACCGCGCATAGTCGCCCGCAAGAGATACCGACATCGCC 1269
QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 1270 TAC-----GGTGGGTAGGTGTAAGTCTACAAGGCC-----ATCAAG 1308
QY 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGln-----LysAlaThr 333
Db 1309 TACGACGACTACCCCAAGTCTACCTTCCCGCTCGCTCGCCGCAAGGAAGCGCCACC 1368
QY 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
Db 1369 CACACCTTCGAGCTCAGCGCGCCACCTTCTGTCGCGCCACCTCTACTGGGAC----- 1422
QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 1423 -----ACGGCTCGAGCGACATCGACTCTTACCTTACGAC 1458
QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnTrpAsp 393
Db 1459 CCCAAC-----GGGAACGAGGTGACTACTCTCTACACCGCTTACTAC--- 1500
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
```

```

Db 1501 -----GGCTTCGAGAGTGGCTACTACACCGCGGCGGAACTGGACGGTC 1551
Qy 414 GluValGlnAlaTyAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGGTCGTCAGCTAC-----AAGGGCGCGGCGGAACTACCGAGTCGACGTCGTC 1599

RESULT 13
ID ABL54900 standard; DNA; 2121 BP.
XX
AC ABL54900;
XX
DT 11-SEP-2003 (revised)
DI 31-MAY-2002 (first entry)
XX
DE T. yonsei subtilisin-like serine protease coding sequence.
XX
KW Subtilisin-like serine protease; ss.
XX
OS Thermococcus aerobacter yonseiensis.
XX
FH Key Location/Qualifiers
FT 142..1179
FT /*tag= a
FT /product= "subtilisin-like serine protease"
XX
KR2000072141-A.
XX
PD 05-DEC-2000.
XX
PF 04-AUG-2000; 2000KR-00045411.
XX
PR 04-AUG-2000; 2000KR-00045411.
XX
PA (KIMV/) KIM Y S.
XX
PI Chang HJ, Kim DH, Byun YR, Kim YS;
XX
DR WPI; 2001-298092/31.
DR P-PSDB; ABB09483.
XX
XX
XX New DNA sequence of thermophilic protein decomposition enzyme and protein
XX derived therefrom.
XX
XX Claim 1; Page 6; 15pp; Korean.
XX
XX This sequence represents the DNA encoding the Thermococcus aerobacter yonsei
XX subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX to standardise OS field)
XX
XX SQ Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;

Alignment Scores:
Pred. No.: 6.68e-20 Length: 2121
Score: 370.50 Matches: 131
Percent Similarity: 45.85% Conservative: 57
Best Local Similarity: 31.95% Mismatches: 139
Query Match: 16.37% Indels: 83
DB: 4 Gaps: 18

US-09-985-689a-3 (1-433) x ABL54900 (1-2121)
Qy 6 GlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyVal 25
Db 445 GGAATCACAAA-----GCACGAGTGAATTTGGAGTCACAGGAAAAAATAAACA 495
Qy 26 ValIleValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAla 45
Db 496 ATACGAATATTGACACAGGATAGCGGAAATCACGTTCACTCTCTCA----- 543
Qy 46 PheArgGlyIleIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63

```

```

Db 544 ---GGTGGAAAAATA-----ATAGATGGAAAGACTTTATCAACAACAA 585
Qy 64 -----AspProAsnGlyHisGlyThrHisVal-----AlaGlySer 75
Db 586 ACTACACCATACGACGACATGGCCATGGAATCTCAGTAGCAAGTATTCTCGAGGTACA 645
Qy 76 ValLeuGlyAsnAlaLeuAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 95
Db 646 GGTGCTGGAAACAGCTCTTTACAAAGCGTTGCTCTGATGCTTTGTTGTAGATAAATA 705
Qy 96 IleMetAspSerSerGly-----GlyLeuGlyGlyLeu 106
Db 706 GTTTTAGATGCAATGGAAGCGGCACATGAGCAGCTGTAACCTGCAGGAATTGACTGGCT 765
Qy 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThr 126
Db 766 GTTCAAAATAAAGATGATATACGGAATCAAAAGTTATAAATTAAGCTCGCCACTTCTACA 925
Qy 127 AsnSerTrpGlyAlaProValaAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGlu 146
Db 826 AGTCTGATGGA-----ACTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 876
Qy 147 TyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSer 166
Db 877 -----ACCGGTATTGTAGTAGTTGTAGCAGCAGGAAACTCTGCGCCCTGCAAAA 924
Qy 167 GlyThrIleSerAlaProGlyThrAlaIysAsnAlaIleThrValGlyAlaThrGluAsn 186
Db 925 TACACCATAGGTCCTGCTGCTGCGGAAAGAACCATACATCCGCGCCCGAGGATATATATA 984
Qy 187 TyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSer 206
Db 985 -----GTAGGTGAACCTTGGCTTTAAC-----CTTCAAGCTTTTCACG 1023
Qy 207 ArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGlyThrPheIle 226
Db 1024 CGCGTCTTACTGCTGACGAGATAAAACCTTCACATTCGCGCCCGAGGATATATATA 1083
Qy 227 LeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnTyrAsnSerLys 246
Db 1084 ACTGCGCGCAAG-----CGCAATTCGTAAATGGA 1113
Qy 247 TyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGln 266
Db 1114 TATGTAACATACACGCGGTACAGCATGGCAACACCTTTGTAGCAGGAACCTGTGTCTT 1173
Qy 267 LeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuIleLysAla 286
Db 1174 ATGCTTAAC-----GCTAATCCAATCTCCTCAATGATGCA-----AAAAAT 1218
Qy 287 AlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTrp 306
Db 1219 ATAATAATGTCTACTGCAAAAAGCTGGGCGCCTCCAAGCAAAACGTTGACTATGGTGCA 1278
Qy 307 GlyArgValThrLeuAspLysSerLeuAsnValAla----- 318
Db 1279 GGAAGATTGGATGCTATGAGCTATAGGCTAGCAGGTAAATTTAGAGGAAATAATATT 1338
Qy 319 -----TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThr 333
Db 1339 GATGTACCAATCATTTATATATTTTCAAGTTATCTGCCCGGTTCTCTCTACAGTGACT 1398
Qy 334 TyrSerPheGln---ThrGlnAlaGlyProLeuLysIleSerLeuValTyrThrAsp 352
Db 1399 TGGACTTTTAACGCAACAATAACAGTTATCCATTCGCAATAACATTAATTAATTCCTGAC 1458
Qy 353 AlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThr 372
Db 1459 -----TGGGCAAAATTAACAAT-----CCCGATTTTGTATATATACCTCTAT 1497
Qy 373 AlaPro---AsnGlyGlnLysTyrValGly 381
Db 1498 GATCCGGCGCATGGAGGAGAAAGACCCCTGGG 1527

```

RESULT 14

ABX34289
 ID ABX34289 standard; DNA; 135638 BP.
 XX
 AC ABX34289;
 XX
 DT 11-FEB-2003 (first entry)
 XX
 DE S. atroolivaceus leinamycin biosynthesis gene cluster.
 XX
 KW Leinamycin biosynthesis gene cluster; Lnm; open reading frame; ORF;
 KW anti-tumour antibiotic; broad spectrum antimicrobial activity;
 KW Gram-positive; Gram-negative bacteria; chemical modification; metabolite;
 KW apo-carrier protein; holo-carrier protein; tumour; polyketide;
 KW hybrid polypeptide/polyketide metabolite; Lnm production; cytostatic;
 KW gene; ds.
 XX
 OS Streptomyces atroolivaceus.
 XX
 XX WO200277179-A2.
 XX
 XX 03-OCT-2002.
 XX
 XX 22-MAR-2002; 2002WO-US008937.
 XX
 XX 26-MAR-2001; 2001US-0278935P.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 XX Shen B, Cheng Y, Tang G;
 XX
 XX WPI; 2003-018907/01.
 DR P-PSDB; ABU11341, ABU11342, ABU11343, ABU11344, ABU11345, ABU11346,
 DR ABU11347, ABU11348, ABU11349, ABU11350, ABU11351, ABU11352, ABU11353,
 DR ABU11354, ABU11355, ABU11356, ABU11357, ABU11358, ABU11359, ABU11360,
 DR ABU11361, ABU11362, ABU11363, ABU11364, ABU11365, ABU11366, ABU11367,
 DR ABU11368, ABU11369, ABU11370, ABU11371, ABU11372, ABU11373, ABU11374,
 DR ABU11375, ABU11376, ABU11377, ABU11378, ABU11379, ABU11380, ABU11381,
 DR ABU11382, ABU11383, ABU11384, ABU11385, ABU11386, ABU11387, ABU11388,
 DR ABU11389, ABU11390, ABU11391, ABU11392, ABU11393, ABU11394, ABU11395,
 DR ABU11396, ABU11397, ABU11398, ABU11399, ABU11400, ABU11401, ABU11402,
 DR ABU11403, ABU11404, ABU11405, ABU11406, ABU11407, ABU11408, ABU11409,
 DR ABU11410, ABU11411.
 XX
 XX Novel gene cluster responsible for synthesis of leinamycin in
 PT Streptomyces atroolivaceus useful for making various peptide and/or
 PT polyketide, and/or hybrid polypeptide/polyketide metabolites.
 XX
 XX
 PS Claim 6; Page 81-127; 185pp; English.
 XX
 XX The present invention relates to the isolation of the Streptomyces
 CC atroolivaceus leinamycin (Lnm) biosynthesis gene cluster containing 71
 CC open reading frames (ORFs) (ORFs -35 through -1, ORFs LnmA through LnmZ,
 CC and ORFs +1 through +9). Leinamycin is a novel anti-tumour antibiotic
 CC produced by several Streptomyces species. It exhibits broad spectrum
 CC antimicrobial activity against Gram-positive and Gram-negative bacteria,
 CC but not against fungi. The polypeptides encoded by the Lnm biosynthesis
 CC gene cluster ORFs are useful for chemically modifying a molecule in a
 CC host cell. The host cell is a bacterium or eukaryotic cell, including a
 CC mammalian, yeast, plant, fungal, or insect cell. The molecule is an
 CC endogenous metabolite produced by the host cell or exogenously supplied
 CC metabolite, or an amino acid, and the polypeptide is a peptide synthetase
 CC or amino transferase. The polypeptides encoded by the Lnm gene cluster
 CC are useful for converting an apo-carrier protein to a holo-carrier
 CC protein. Lnm shows potent antitumour activity in tumour models in vivo.
 CC The Lnm gene cluster modules and/or catalytic domains are useful for
 CC making various peptide and/or polyketide, and/or hybrid
 CC polypeptide/polyketide metabolites. The proteins encoded by the ORFs are
 CC useful alone, or in combination with other active domains to modify
 CC various target substrates. The Lnm gene cluster is useful to upregulate
 CC endogenous Lnm production to permit Lnm production in cells and/or to

CC make various modified Lnm. Lnm, its analogue, or other polyketide,
 CC peptide or hybrid polyketide/peptide metabolites are useful as
 CC therapeutic agents, to treat a number of disorders, depending upon the
 CC type of metabolites. The present sequence represents the S. atroolivaceus
 CC leinamycin biosynthesis gene cluster
 XX
 SQ Sequence 135638 BP; 18570 A; 49096 C; 49039 G; 18933 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8 34e-16 Length: 135638
 Score: 346.50 Matches: 140
 Percent Similarity: 42.39% Conservative: 55
 Best Local Similarity: 30.43% Mismatches: 154
 Indels: 111
 Query Match: 7
 Gaps: 23
 DB:
 US-09-985-689A-3 (1-433) x ABX34289 (1-135638)
 QY 6 GlyTleVallysaAlaaspValAlaGlnAsnAsn----- 16
 DB 8067 GGCAGGGTGAAGCCGATCTGGCCGACTCCACGCCAGATCGCGCGCAGAGGTATGG 8126
 QY 17 ---TyrGlyLeuTyrGlyGlnGlyGlnValValAlaAlaaspThrGlyLeuAspThr 35
 DB 8127 GCGAGGGCGACACCGCCAGGACGTCAAGTCTGCGATGCTGCACAGCGCGGACACC 8186
 QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysle-----ThrAlaLeu 53
 DB 8187 -----GAAACCCCGAGACTCTGTCGGCAGGTCTCCGACGCGCGCAGC 8228
 QY 54 TyrAlaLeuGlyArgThrAsnAlaAsnAspProAsnGlyHisGlyThrHisValAla 73
 DB 8229 TTGTCCTCCCGCGAGGACGACATCGCC--GATTACACGCCCCACGCGCAGCTCGCC 8285
 QY 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88
 DB 8286 TCGACCATCTCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8345
 QY 89 AlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSer 108
 DB 8346 GCGCGGTGTCGTCGCGCAAGGTGCTCACTCCGAGGC-----AGCGCGCAGGATCG 8399
 QY 109 AsnLeuAsnThrLeuPheSerGlnAlaTTPAsnAla-----GlyAlaArgIleHis 125
 DB 8400 TGGATC-----ATCGCGGCGATCGAGTGGCGCGCGCGCGCGCGCGCGCGCG 8453
 QY 126 ThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn-----SerArg 142
 DB 8454 AGCATGAGCTCGGC-----GGCGCGGTGACCAAGAACCGCGCGCGCGCGCGCG 8501
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 DB 8502 GCGTCGACGAACTCAGCCACGACACGCGCGCGCTTGTTCGTGATCGCGCGCGCGCG 8561
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 DB 8562 GCGCGCACTCC-----ATCAGACGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8615
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
 DB 8616 GCC-----GTCGACTCCACCGACGCTCGCC 8642
 QY 203 GlnPheSerSerArgGly-----AlaThrArgAspGlyArgIleLysProAspValThrAla 221
 DB 8643 GACTTCTCCAGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8702
 QY 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAla 241
 DB 8703 CCGCGCGTCGACATCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8756
 QY 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 261
 DB 8757 -----TACACCGATGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8801


```
QY 262 GlyAsnValAlaGlnLeuArgLHisPheIleLysAsnArgGlyLeuThrProLysPro 281
DB 8802 GGTGTCCCGCGGCTCTCTCCCGCGAGCACCCCGAGTGGACGGCACCAGCTCAAGGAG 8861
QY 282 SerLeuLysAlaAlaLeuLeuAlaGlyAlaThrAspValGlyLeuGlyTyrProSer 301
DB 8862 GCACGTGTCAGCAGCCCAAGCAAGCAGCGCGG-----TACACCCG 8903
QY 302 GlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsn 321
DB 8904 TACCAGCGCGCGCGCGCTCGACGGCGCGCGCTGCACACACACAGCTCTTCGCC 8963
QY 322 GluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAlaGly 341
DB 8964 ACCACAGCGCGCTACTCC-----GGCTTCACACG----- 8993
QY 342 LysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThr-----Thr 358
DB 8994 -----TGGCCCCCGAAGCCCGCGGAGACCGATGTCGGGACG 9029
QY 359 AlaSerTyrThrLeuValAsnAsp-----LeuAspLeuValIle-----Thr 372
DB 9030 GTACGCTTACCAACCTCGCGGACGCGCGCTCAGCTCAACCTGCGCTCAACGGCAC 9089
QY 373 AlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrp 392
DB 9090 GTCCCGCGCGGTG-----TTCAGGCTCTCCGAGGACCAT----- 9125
QY 393 AspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThr 412
DB 9126 -----GTACCGTGCCTCCGCGCACCGGACCGCGCCACG 9155
QY 413 IleGluVal--GlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
DB 9156 GTCACCTGACCGCGGTCTTGACACAGCTGGCGGCGGACCGTCCGTACGCGCGGTGATC 9215

RESULT 15
AAAT61454
ID AAAT61454 standard; DNA; 2539 BP.
XX
XX
AC AAAT61454;
DT 06-OCT-1997 (first entry)
DE
KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
XX
OS Streptomyces viridosporus.
FH Key Location/Qualifiers
FT CDS 338..2539
FT /tag= a
FT /note= "no stop codon given"
FT misc_feature 950..2509
FT /tag= b
FT /note= "encodes AAW13666"
XX
PN WO9705243-A1.
XX
PD 13-FEB-1997.
XX
PF 30-JUL-1996; 96WO-JP002147.
XX
XX 31-JUL-1995; 95JP-00212975.
PR 29-FEB-1996; 96JP-00067478.
XX
PA (SAOC ) MERCIAN CORP.
XX
PI Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;
```

```
PI Isshiki K, Yoshioka T;
XX
XX WPI; 1997-145682/13.
DR P-PSDB; AAW13666, AAW13667.
XX
PT Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
PT on 4-substituted-1,4-di:hydro:pyridine derivatives to produce chiral
PT derivatives useful for synthesis of cardiovascular drugs.
XX
XX Claim 3; Page 49-55; 78pp; Japanese.
XX
XX This sequence is the Streptomyces viridosporus dhpA gene which encodes an
CC asymmetric hydrolase which acts on 4-substituted-1,4-dihydropyridine
CC derivatives. The enzyme allows the efficient conversion of 4-substituted-
CC 1,4-dihydropyridine esters to chiral partially hydrolysed derivatives,
CC for use in the synthesis of cardiovascular drugs suitable for the
CC treatment of e.g. hypertension and ischaemic heart disease
XX
XX Sequence 2539 BP; 433 A; 959 C; 867 G; 280 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 1.15e-17 Length: 2539
Score: 343.50 Matches: 124
Percent Similarity: 38.76% Conservative: 45
Best Local Similarity: 28.44% Mismatches: 150
Query Match: 15.18% Indels: 117
DB: 2 Gaps: 17
US-09-985-689A-3 (1-433) x AAT61454 (1-2539)
QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
DB 998 GGCTACGACGCAAGGGGTGAAGATCGCCTCTCGACACCGGTGTCGACAGCAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
DB 1055 -----CATCCGACCTGAAGGGCGGGTGACCGGCTCAAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValleu 77
DB 1100 GCCGCGCGCGCGCGCGGACAGAGTGCGCCACCGCACCGCTCGCTCGATCGCGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
DB 1160 GGCACGGGCGCCAGTCCACAGGCAAGTACAAAGGGCTGCGACCGCGCGCGCTCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyLeuGlyLeuProSerAsnLeuAsnThr 112
DB 1220 AACGGCAAGGTCTCTGACGACTCC-----GGTTTCGGGACGACTCCGGCATC 1267
QY 113 LeuPheSerGlnAlaTyrAsnAla-----GlyAlaArgIleHisThrAsnSerTyrGly 130
DB 1268 CTCGCGCGCATGGAGTGGCGCGCGCGCGCGCGCGCTCGTCAACATGAGCTGGGC 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgln 143
DB 1328 GGCATGGACACACCGGAGACCGACCCGTGGAGCGCGG----- 1366
QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
DB 1367 GTCGACAGCTGTCGCGGAGAGGGCGTCTGTTTCGCTCGCGCGCGCGCAACGAGGCG 1426
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
DB 1427 CCGGAGTCG-----ATCGGTTTCGCGCGGACCGCGCGCGCTCACGTCGCGCGCC 1480
QY 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
DB 1481 -----GTCGACGACAGGACAGCAAGCTCGCCGAC 1507
QY 204 PheSerSerArgGly-----AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
DB 1508 TTCCTCTCACCGCGCGCGCGCTCGCGGACCGCGCGCATCAAGCGGAGCTCACCGCTCCC 1567
```

223	QY	GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerThrTrpAlaAsn	242
1568	Db	GGCGTGGACATCAGCGCGGCTCGGGGAGGCAACGATCGCCGAGAGTGGGTGAG	1627
243	QY	TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly	262
1628	Db	GGACGGCGGGTACATGACCATCTCGGCACGTCATGGCACCCTCGACGTCGGGGC	1687
263	QY	AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgLysIleThrProLysProSer	282
1688	Db	GGGGCGGCGCTCTCGAAGCAGCAG-----CACCCCGAC	1720
283	QY	LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr	299
1721	Db	TGGACCTCGCGCAACTGAAGGGCGGCTCACCGGCTCCACCAAGGGCGGC---AAATAC	1777
300	QY	ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysAserLeuAsnValAlaTyr	319
1778	Db	ACCCGCTTCGAGCAGAGGGTTCGGGCGGATCCAGGCCGACAGGGCGTCCAG-----	1828
320	QY	ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln	339
1829	Db	-----CAGACCCGT	1837
340	QY	AlaGlyLysProLeuLysIleSerLeu-----ValTrp-----ThrAspAla	353
1838	Db	ATCGCGACCCGCTCCGTAGCTTCGGCGTCCAGTGGCGGCACACCGACGACGAG	1897
354	QY	ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle-----	371
1898	Db	CGGTCACCAAGCAGGTGACCTACCGCAACCTCGGCACCCGACGACGTACGCTGAAGCTG	1957
372	QY	-----ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrPro	387
1958	Db	ACGTCGACCGCCACCCGACCCCAAGGGCAAG-----	1987
388	QY	TyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln	407
1988	Db	-----GGGGCCCCG	1996
408	QY	SerGlyThrTyrThrThrIleGluValGlnAlaTyrAsnValProSerGly	423
1997	Db	GGGGGCTTCTTACGCTGGGGCGCCACACCGGTACCGTCCCGCGGGC	2044

Search completed: April 4, 2004, 01:11:31
Job time : 392.908 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.4512 Seconds
(without alignments)
3787.066 Million cell updates/sec

Title: US-09-985-689A-3
Perfect score: 2263
Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVPSGQRFSLAIVH 433

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 segs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09985689/runat_31032004_161807_4184/app.query.fasta_1.3498
-DB=issued_patents.NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -NCPUS=3
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents.NA:
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCBUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2040	90.1	3003	2	US-08-873-479-41
2	1998.5	88.3	1920	4	US-09-509-814A-3
3	1995.5	88.2	1923	4	US-09-509-814A-7
4	1994.5	88.1	1923	4	US-09-509-814A-5
5	443	19.6	1977	3	US-08-894-818B-2
6	443	19.6	1977	4	US-09-445-472-11
7	422.5	18.7	1236	3	US-09-445-472-2
8	422.5	18.7	1566	3	US-08-894-818B-4
9	422.5	18.7	1962	3	US-08-894-818B-34
10	422.5	18.7	1962	4	US-09-445-472-15
11	401.5	17.7	1977	3	US-08-894-818B-6
12	348.5	15.4	2539	3	US-09-000-016-3

13	348.5	15.4	2539	4	US-09-514-340-3
14	343.5	15.2	2809	3	US-09-000-016-1
15	343.5	15.2	2809	4	US-09-514-340-1
16	312	13.8	2532	1	US-07-671-3760-4
17	306.5	13.5	1859	3	US-08-894-818B-15
18	305	13.5	2835	1	US-08-750-532-2
19	305	13.5	4765	1	US-08-750-532-8
20	305	13.5	4765	3	US-08-894-818B-7
21	305	13.5	4765	4	US-09-445-472-5
22	282	12.5	564	1	US-08-750-532-11
23	282	12.5	564	3	US-08-894-818B-14
24	281.5	12.4	898	1	US-08-750-532-7
25	269	11.9	1140	1	US-08-322-677A-11
26	269	11.9	1140	1	US-08-322-676-11
27	269	11.9	1140	3	US-08-898-218-11
28	269	11.9	1140	3	US-08-848-793-11
29	269	11.9	1140	4	US-09-445-270-6
30	269	11.9	1140	4	US-08-322-678-11
31	269	11.9	1143	3	US-08-269-050-1
32	269	11.9	1143	6	5336611-1
33	269	11.9	1320	6	5217878-1
34	269	11.9	1350	3	US-09-049-867-1
35	269	11.9	2280	1	US-07-661-378A-1
36	269	11.9	2280	1	US-07-918-318-1
37	269	11.9	2280	2	US-08-413-724-1
38	269	11.9	2280	3	US-08-853-494-1
39	269	11.9	10216	2	US-08-875-154-1
40	268	11.8	807	1	US-07-706-691C-72
41	268	11.8	807	1	US-08-254-021-72
42	268	11.8	807	2	US-08-618-446-72
43	268	11.8	807	3	US-08-980-135-72
44	268	11.8	807	4	US-09-585-798-72
45	268	11.8	1140	1	US-08-322-677A-12

ALIGNMENTS

RESULT 1

US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lynne, Christianson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agilis, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:

```
/ LENGTH: 3003 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 5,066-194 Length: 3003
Score: 2040.00 Matches: 385
Percent Similarity: 94.46% Conservative: 24
Best Local Similarity: 88.91% Mismatches: 24
Query Match: 90.15% Indels: 0
DB: 2 Gaps: 0

US-09-985-689A-3 (1-433) x US-08-873-479-41 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCGTGGCATTGTGAAGCAGACGTCGCACAAATAAATCTTTGGCTTATAT 1529

Qy 21 GlyGlnGlyGlnValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GCACAGGACAGATGTAGCAGTTCCTGATCTGGCTTGATACAGGAGAAATACACAGT 1589

Qy 41 SerMetHisGluAlaPheArgGlyIysIleThrAlaLeuTyrAlaLeuGlyValArgThrAsn 60
Db 1590 TCGATGCATGACGATTCGCGGTGAAGATTACGACATATATGCTGGCAGAGCAAT 1649

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCAATGTGCTGGATCTGTGTAGGAATGCT 1709

Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
Db 1710 ACAATTAAGGATGACGCGGACCAATCTAGTCTTTCATCTATATATGATAGTGGT 1769

Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrAsnAla 120
Db 1770 GGAGGGCTGGAGGACCTACCTGCTAATCTACAAACATATTATTCAGTCAAGCATATAGTGT 1829

Qy 121 GlyAlaArgIleHisThrAsnSerTTPGlyAlaProValAsnGlyValAlaTyrThrAlaAsn 140
Db 1830 GGAGCGAGATTCATAGCAATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACAGAC 1889

Qy 141 SerArgGlnValAspGlnTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
Db 1890 TCTCGAATCTTGATGATTATGTGAGAAAAATGATATGACGATCTTTTTCGGCGCGGA 1949

Qy 161 AsnGlyGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1950 AATGAGGACCGAGTACGCTGATCANTAGTGCACCGAGAACAGCAAAAATGCGATTACA 2009

Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 2010 GTTGGGCAACCGAAACCTACGTCCAAAGCTTCGGATCTTATCGGATAATATTAACCAT 2069

Qy 201 IleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 2070 GTTCTCAATCTCTTCACGAGGTCCTACTAGAGATGGAGTATTAAAGCCGCGCATG 2129

Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
Db 2130 GCACCGAGTACGTATATCTCTGCTAGATCATCATAGTCTCCAGATTCCTCATCTGG 2189

Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 2190 GCAAAACCATAGATTAATATGCTACATGGGTGATCTTCTATGGCTACTCCAAATGTA 2249

Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
Db 2250 GCAGGTAATGTGGACAAATTAAGGAGCATTTTGTGAAAATAGAGGGGTAACTCTCTAAG 2309
```

```
Qy 281 ProSerIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
Db 2310 CTTTCCTTTTAAAGCTGCTTTAATGTCAGGTGTCGGGATGTTGGACTTGCCTTCCA 2369

Qy 301 SerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
Db 2370 AATGGTAACCAAGGATGGGAGAGTAACGTTAGATAAATCCCTAAATGTCGCAATTTGT 2429

Qy 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
Db 2430 AATGAAACGAGCCCTTTATCAACAGTCAAAAGAACATATTCGTTTACGGCTCAAGCT 2489

Qy 341 GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer 360
Db 2490 GGTAAACCTTAAATAATACACTTGTGTCAGATGCACAGGTAGCAGCGGCATCA 2549

Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
Db 2550 CTAACCTTTAGTGAATGATTAGACTTTAGTAATCACTGCACCAATGGAATTAATACGTC 2609

Qy 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsn 400
Db 2610 GGAATGACCTTACAGCACCGTATGATAACAATGGATGGCAGAAACAAACGTTGAAAT 2669

Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
Db 2670 GTGTTTATCAATGCTCTCCTCAAGCGGAACGTTATACAGTCAAGTGCAGGCTTACAATGTA 2729

Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 2730 CCAGTAAGTCCGCAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 2
US-09-509-814A-3
Sequence 3, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAEKI, KATSUHIISA
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUW
APPLICANT: NOMURA, WASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-0PCT
CURRENT APPLICATION NUMBER: US/09/509,814A
CURRENT FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JP98/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
PRIOR FILING DATE: 1997-06-08
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patent in version 3.0
SEQ ID NO 3
LENGTH: 1920
TYPE: DNA
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1920)
US-09-509-814A-3

Alignment Scores:
Pred. No.: 3,616-190 Length: 1920
Score: 1998.50 Matches: 382
Percent Similarity: 94.01% Conservative: 26
Best Local Similarity: 88.02% Mismatches: 25
Query Match: 88.31% Indels: 1
DB: 4 Gaps: 1

US-09-985-689A-3 (1-433) x US-09-509-814A-3 (1-1920)
```


QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGGATGAGGACCGAAGCGGAGAACCATCATAGTCACCGAGCACAGTAAATATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACATGCGAGCTACGGAACCTCGGCCAAGCTTGGGTCTTATGCGGACATATCAAC 1218
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGGCACAGTCTCTTTCAGTGGACCGCAAGAGATGGACGATCAACCGGATGTC 1278
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGACCGGACGTTCTACTATCAGCAAGATCTTCTTGACCGGATCTCTCTTC 1338
QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1339 TGGGCGAACCATGACAGTAATATGATACATGATGGTGGAGCTCATGGCTACACCGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1399 GTTGCTGGAACGTGGCAGCTTCTGAGCATTTTGTGAACAGAGGATCACACCA 1458
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCTTCTCTATTAAAGCGGCACTGATTCGCGGTGACGCTGACATCGGCTTGGCTAC 1518
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrIleAspLysSerLeuAsnValAlaTyr 319
Db 1519 CGAAGCGTAACCAAGATGGGAGGAGTGGATGATGATGATGATGATGATGATGATGAT 1578
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1579 GTGACGAGTCCAGTCTCTATCCACGAGCAAAAGGACGATCTGTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrThrAla 359
Db 1639 GCGGCAAGCCTTGAATATCCTCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1698
QY 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCGTAACGCTTGTCAATGATCTGACCTTGTCTATACCGCTCCCAATGGCACAGTAT 1758
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGlu 399
Db 1759 GTAGGAATGACTTTTCTTGGCATAATGATGATGATGATGATGATGATGATGATGATGAT 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGATTTATTATGACCAACAAGCGGAGCTATACATTTAGGTACAGGCTTATTAAC 1878
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGTTGGACACAGACCTTCTCTGTTGCAATTTGTAAT 1920

RESULT 5
US-08-894-818B-2
Sequence 2, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:
Pred. No.: 1,43E-34 Length: 1977
Score: 443.00 Matches: 137
Percent Similarity: 46.03% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 170
Query Match: 19.58% Indels: 68
DB: 3 Gaps: 16

US-09-985-689A-3 (1-433) x US-08-894-818B-2 (1-1977)
QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyGlnValValAla 27
Db 433 ATAGGGGCGGATACCGTCTGGAACCTCCTCGGCTACGACGAAGCGTGTGGTGTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
Db 493 ATGTCGATACGGGTATAGCGGAAC-----CACCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
Db 535 GGCAGGTCATAGGCTGTGACGACGCGCTCAACGCGAGGTCTGACCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
Db 595 GGCACGGAACCCACCGTTGGGTATCGTTCGCGAACCGGACGCGGCTTAACTCCAGTAC 654
QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
Db 655 ATAGGCGTCCCGCGCGGAAGCTCGTGGCGGTCAAGGTTCTCGGTGCGGACGCTTCG 714
QY 101 GlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 715 GGAAGCGTCTCCACCATCATCGCGGGGTGTGACTGGTCTGTCGACGACGACGACGACG 774
QY 121 GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 775 GGGATAGGTCATCAACCTCTCCTCGGCTCTCTCCAGAGCTCCGACGAGACCGACTCC 834

APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA=6
CURRENT APPLICATION NUMBER: US/09/445,472
CURRENT FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in version 3.0
SEQ ID NO 11
LENGTH: 1977
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-09-445-472-11

Alignment Scores:
Pred. No.: 1,43e-34 Length: 1977
Score: 443.00 Matches: 137
Percent Similarity: 46.03% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 170
Query Match: 19.58% Indels: 68
DB: 4 Gaps: 16

US-09-985-689a-3 (1-433) x US-09-445-472-11 (1-1977)

QY 8 VallysAlaAspValAlaGlnAsnAsnTyxGlyLeuTyxGlyGlnValValAla 27
Db 433 ATAGGGCGGATACCGCTCGGAACCTCCCTCGCTACCGAAGCGGTGGTGGTGGC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCTCGATACGGGTATAGACCGAAC-----CACCCGATCTGAAG 534
QY 48 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
Db 535 GGCAGAGTCATAGCTGTCGACGCGCTAACCGCAGGTGACCCCTACGATGACGAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
Db 595 GGACACGAGAACCCAGCTTCGGGTATCGTTCGCGAACCGCAGCGGTAACTCCAGTAC 654
QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
Db 655 ATAGCGTCGCGCGCGCGGAGCTCTCGCTCGCTCAGAGTTCCTCGGTGCGCGAGTTCG 714
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTGACCTGGTTCGTCGACAGCAAGCAAGTAC 774
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsn 140
Db 775 GGGATAAGGGTTCATCAACCTCTCCCTCGCTCTCCAGAGCTCCGACGGAACCGACTCC 834
QY 141 SerArgGlnValAspGluTyxValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
Db 835 CTCAGTCAGCGCGCTCAACACCGCTGGGACCGCGGTATAGTAGTCTGGTGGCGCGCGC 894
QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 895 AACAGCGGCGCGGACACACCTACCTCGCTCACCCTCGCTCACCCTCGCTCAGCAAGGTCTAACC 954
QY 181 ValGlyAlaThrGluAsnTyxArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 955 GTGCGTGCA-----GTTGACAGCAACGACACAC 981
QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 982 ATGCCAGCTTCTCCAGCAGGAGGACCGACCGCGGACGGAAGGCTCAAGCGGAAGTCGTC 1041
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPheTrp 240

QY 141 SerArgGlnValAspGluTyxValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
Db 835 CTCAGTCAGCGCGCTCAACACCGCTGGGACCGCGGTATAGTAGTCTGGTGGCGCGCGC 894
QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 895 AACAGCGGCGCGGACACACCTACCTCGCTCACCCTCGCTCACCCTCGCTCAGCAAGGTCTAACC 954
QY 181 ValGlyAlaThrGluAsnTyxArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 955 GTGCGTGCA-----GTTGACAGCAACGACACAC 981
QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 982 ATGCCAGCTTCTCCAGCAGGAGGACCGACCGCGGACGGAAGGCTCAAGCGGAAGTCGTC 1041
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPheTrp 240
Db 1042 GCCCGCGCGCTTACATCATAGCCCGCGCGCGCAGC-----GGAACGATGAGCGC 1092
QY 241 AlaAsnTyxAsnSerIleAlaTyxMetGlyGlyThrSerMetAlaThrProIleVal 260
Db 1093 ACCCGGATACGACTACTACCAAGGCTCTGGAACGAGCATGCGCCACCGCGACGTT 1152
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleIleIleAsnArgGlyIleThrPro--- 279
Db 1153 TCGGGCGCTTCGCGGCTCATCTCTCCAG-----GCCACCGAGCTGGACCCCGGAC 1203
QY 280 -----LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
Db 1204 AAGGTGAAGACCGCGCTCATCGAGCGCGGACATAGTCGCCCCCAAGGAGATAGCGGAC 1263
QY 297 LeuGlyTyxProSerGlyAspGlnGlyTyxGlyArgValThrLeuAspLysSerLeuAsn 316
Db 1264 ATCGCTTAC-----GGTGGGTAGGTGAACGTCTACAGGCC-----Lys 331
QY 317 ValAlaTyxValAsnGluAlaThrAlaLeuThrThrGlyGln-----Lys 331
Db 1303 ATCAAGTACGACGACTACGCAACCTCACCTTCACCGCTCCGTCGCGCAAGGGAAGC 1362
QY 332 AlaThrTyxSerPheGlnThrGlnAlaGlyLysProLeuIleLysLeuValThrTrp 351
Db 1363 GCCACCCACACCTTCGAGCTCAGCGGCGGCGCACCTTCGTCGCGCCACCTCTACTGGGAC 1422
QY 352 AspAlaProGlySerThrThrAlaSerTyxThrLeuValAsnAspLeuValIle 371
Db 1423 -----ACGGCTCGAGCGACATCGACCTCTACCTC 1452
QY 372 ThrAlaProAsnGlyGlnLysTyxValGlyAsnAspPheSerTyxProTyxAspAsnAsn 391
Db 1453 TACGACCCCAAC-----GGGAACGAGGTGACTACTCTCTACACCGCGCTAC 1497
QY 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyx 411
Db 1498 TAC-----GGTTCGAGAAGTTCGGCTACTACACCGCGCGGAGGAGTCTG 1545
QY 412 ThrIleGluValGlnAlaTyxAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
Db 1546 ACGGTCAAGTTCGTAGCTAC-----AAGGCGCGCGGAGGAGTCTACAGGTCGACGTC 1596
QY 432 Val 432
Db 1597 GTC 1599

RESULT 6
US-09-445-472-11
; Sequence 11, Application US/09445472
; Patent No. 6356728
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHINOJO, Tomoko

Db 1042 GCGCCGGGCTGACATCATAGCCCGCGGCAGC-----GGACACGATGGGC 1092
Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 1093 ACCCGGATAAAGCACTACTACACCAAGGCTCTCGAACCAGCATGGCCACCCGACGTT 1152
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
Db 1153 TCGGGGCTGGCGGCTCATCTCCAG-----GCCACCCGAGTGGACCCCGGAC 1203
Qy 280 -----LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGly--- 296
Db 1204 AAGGTGAAGACCGCCCTCATCGAGCCGCGACATAGTCGCGCCCAAGAGATAGCGGAC 1263
Qy 297 LeuGlyTyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsn 316
Db 1264 ATCGCCTAC-----GGTGGGTAGGTGAAGCTTTCACAGGCC----- 1302
Qy 317 ValAlaTyrValAsnGluAlaThrAlaLeuThrThrGlyGln-----Lys 331
Db 1303 ATCAAGTACGACGACTACGCCAAGCTACCTTCACCGGCTCGCGGACCAAGGAAGC 1362
Qy 332 AlaThrTyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThr 351
Db 1363 GCCACCCACACCTTCGACGTACGCGCGCCACCTTCGTGACCGCCACCCCTCTACTGGGAC 1422
Qy 352 AspAlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuValIle 371
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTC 1452
Qy 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
Db 1453 TAGACCCCCAAC-----GGGAACGAGGTTCGACTACTCTACACCCCGCTAC 1497
Qy 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
Db 1498 TAC-----GGCTTCGAGAGGTTCGCTACTACACCCCGCGGACCTGG 1545
Qy 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
Db 1546 ACGGTCAAGGTCTCGACTAC-----AAGGGCGCGCGGAACCTTACCAGGTCTGACGTC 1596
Qy 432 Val 432
Db 1597 GTC 1599

RESULT 7

US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445, 472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2

Alignment Scores: 7.84e-33 Length: 1236
Pred. No.: 422,50 Matches: 138
Score: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.67% Indels: 79
DB: 4 Gaps: 17
US-09-985-689A-3 (1-433) x US-09-445-472-2 (1-1236)
Qy 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTGGGATATGATGATGCTTCTGAATACACATAGGAATTAATGACACTGGAATGAC--- 114
Qy 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 115 -----GCTTCTCATCCAGATCTCCAAGGAAAGTA----- 144
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATTGGGGTGGTATGATTTGTCATGTGTAGGATTAATCCATGATGACCATGACATGGA 204
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 205 ACTCATGTAGCTCAATAGCAGCTGTTACTGGAGCAGCAAGTAATGCAAGTACAAAGGA 264
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 265 ATGCTCTCAGGAGTACAGTGGCGGGAATTAAGGTTCTAGGTCCGATGTTCTGGAGC 324
Qy 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 325 ATATCTACTAATATTAGGAGTTGAGTGGCGCTTGCATTAACAAGATAAGTACGGAATT 384
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 385 AAGGTCAATTAATCTTCTCTGTTCAAGCCAGAGCTCAGATGCTACTGACGCTCTAAGT 444
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTTAATGAGCGGTGGATGCTGGATTAGTTGTTGTTGCTGCCGTGGAACAGT 504
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTAACAGTATACAAATCGGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGA 564
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 565 GCCGTTGACCAAGTAT-----GATGTTATACA 591
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 592 AGCTTCTCAAGCAGAGGGCCAACTGCGACGCGAGGCTTAAGCTGAGGTGTTGCTCCA 651
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 652 GGAACCTGGATTAATGCTGCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 702
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACACAGCAGCTCTGCGGACATCAATGCAACTCTCCTCAGTGTGTT 762
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 763 ATTGAGCCCTCTTGTCTCCAA-----GCACACCCGAGCTGGACTCCAGACAAAGTA 813
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 814 AAAACAGCCCTCATAGAACTGCTGATATCGTAAGCCAGATGAATAGCCGATATAGCC 873
Qy 299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 874 TAC-----GGTGCAGGTAGGTTAATGCATACAGGCTATAAAC----- 912

QY 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
Db 913 TAGGATACTATCAAGAGTAGTGTCTACTGGATATGTTGCCAACAAAGGCGCAACT 972
QY 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThrThrAspAla 353
Db 973 CACCACTTCCTGATTAGCGAGCTTCGTTCACTGCAATGCGCACTTATCTGGGACCAATGCC 1032
QY 354 ProGlySerThrThrAlaSerThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 1033 AAT-----AGCGACTTGATCTTACTCTACGAT 1062
QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db 1063 CCCAATGGAAACCAAGTT-----GACTACTCTTAC-----ACC 1095
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
Db 1096 GCTACTATGATTCGAAAGTTGGTTATTATCAACCCCACTGATGGAACATGGCAATT 1155
QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1156 AAGTTTGAAGCTAC-----AGCGGAAGTGCAAACTATCAAGTAGATGTGTA 1203

RESULT 8

US-08-894-818B-4
; Sequence 4, Application US/0894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; OTHER INFORMATION: /note= N at position 1283 is G or T.
US-08-894-818B-4
Alignment Scores:
Pred. No.: 1,13e-32 Length: 1566
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.67% Indels: 79
DB: 3 Gaps: 17

US-09-985-689A-3 (1-433) x US-08-894-818B-4 (1-1566)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACITGGGATATGATGTTCTTCTGAATCAATAGGAATAATGACACTGGAATTGAC--- 114
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 115 -----GCTTCTCATCCAGATCTCCAAGGAAAAGTA----- 144
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATGGGTGGGTAGATTTTGTCAATGCTAGGAGTTATCCATACCATGACCATGGCATGGA 204
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGCAAGTACACAGGA 264
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 265 ATGGCTCCAGGAGCTAAGCTGCGGAATTAAAGTTCTAGGTGCGGATGTTCTGGAGGC 324
QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 325 ATATCTACTATAATAATTAAGGGAGTTGAGTGGCCCTTGATTAACAAGATAAGTACCGA 384
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 385 AAGGTCAATTAATCTTCTTGGTTCAAGCCAGAGCTCAGATGTTACTGACGCTCTAAGT 444
QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTTAATGACGCGTGGATGCTGATTGTTGTGTTGGCTGGAAACAGT 504
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTACAAGTATACATCGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGGA 564
QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 565 GCCGTTGACAGTAT-----GGAACCTAGCATCGGTCAACCA 591
QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 592 AGCTTCTCAAGCAGAGGGCCCACTCCAGCAGGAGCTTAAGCCCTGAGTTGTTGCTCCA 651
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 652 CGAAACTCGATAATTGCTGCCAGAGCAAGT-----GGAACCTAGCATCGGTCAACCA 702
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACAGCAGCTCTCTGGGACATCAATGGCACTCTCTCAGTAGCTGT 762
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 763 ATTGAGCCCTCTTGTCTCAA-----GCACACCGAGCTGGACTCCAGCAAAAGTA 813
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298

```
Db 814 AAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATAGCCGATATAGCC 873
Qy 299 TyrProSerGlyAspGlnGlyTyrPheGlyValThrLeuAspLysSerLeuAsnValAla 318
Db 874 TAC-----GGTGCAGTAGGGTTAATGATACAGGCTATTAAC----- 912
Qy 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
Db 913 TACGATAACTATGCAAGCTAGTGTCTACTGGATATGTTGCCAACAAGGCGACCAAACT 972
Qy 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTyrThrAspAla 353
Db 973 CACCAGTTCTGTTATTAGCGAGCTTCGTTGTAACCTGCCATATATCTGGCAAACTGCC 1032
Qy 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuValIleThrAla 373
Db 1033 AAT-----ACGCGACCTTGATCTTTTACCTCTACGAT 1062
Qy 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrPasp 393
Db 1063 CCCAATGAAACCCAGGTT-----GACTACTCTTAC-----ACC 1095
Qy 394 GlyAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
Db 1096 GCCTACTATGGATTGAAAGGTTGGTTATTACACCCCACTGATGGACATGGACAAATT 1155
Qy 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1156 AAGGTTGTAAGCTAC-----ACGCGAAGTGCAAACTATCAAGTAGATGTGTA 1203
```

RESULT 9

US-08-894-818B-34
Sequence 34, Application US/088994818B
Patent No. 6261822

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITA, Masanori
APPLICANT: ASADA, Kiyoko
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunobu
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/894,818B

FILING DATE: 20-MAY-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03253

FILING DATE: 07-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 323285/1995

FILING DATE: 12-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TAKAKURA=1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1962 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-34

Alignment Scores:

Pred. No.: 159e-32 Length: 1962
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.67% Indels: 79
DB: 3 Gaps: 17

US-09-985-689A-3 (1-433) x US-08-894-818B-34 (1-1962)

```
Qy 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTGGGATATGATGTTCTGGAATCACAATAGGAATTAATGACACTGGAATTGAC--- 510
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----CCTTCTCATCCAGATCTCCAGAGAAAGTA----- 540
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATTGGGTGGGTAGATTGTCATATGTTAGGAGTTATCCATAGCATGACCATGGACATGGA 600
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 601 ACTCATGTAGCTTCAATAGCAGCTGTGTACTGGAGCAGCAAGTAATGGCAAGTACAGGGA 660
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 661 ATGGCTCCAGAGCTAAGCTAGTGGCGGAATTAAGGTTCTAGTCCGCGATGTTCTTGAAGC 720
Qy 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 721 ATATCTACTATAATAAGGAGTTGAGTGGCGCGTGTGATACCAAGATAAGTACGGAAAT 780
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValAlaTyrThrAlaAsnSerArg 142
Db 781 AAGGTCATTAACTCTTCTCTGTTTCAAGCCAGAGCTCAGATGTTACTGACGCTCTAAGT 840
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaGlyAsnGlu 162
Db 841 CAGGCTGTTTAATGACGCTGGGATGCTGGATTAGTTGTTGTTGCTGCCCTGGAAACAGT 900
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 901 GGACCTAACAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGATTATTACAGTTGA 960
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 961 GCGGTTTCAACAAGTAT-----GATGTTATAACA 987
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 988 AGCTTCTCAAGCAGAGGCGCAACTGCAGCGGAGGCTTAAGCTGAGGTTGTTGTTCCA 1047
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1048 GGAACCTGGATAATTGTCGCAGAGCAAGT-----GGAACTAGCATGGGTCAACCA 1098
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1099 ATTAATGACTATTACAGCAGCTCTGGGACATCAATGGCAACTCTCTACGTTAGTGGT 1158
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
```


QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnTrpAsp 393
 Db 1459 CCCATGGAACAGGTT-----GACTACTTTTAC-----ACC 1491
 QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
 Db 1492 GCCTACTATGATTGCAAAAGGTTGTTATTACAAACCACTGATGGAAACATGGACAATT 1551
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1552 AGGTTGTAAAGTAC-----ACCGAAGTGCMAAATATCAAGTAGATGGTGA 1599

RESULT 11

US-08-894-818B-6
 ; Sequence 6, Application US/08894818B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MITTA, Masanori
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; STREET: Browdy and Neimark
 ; ADDRESS: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: 08/08/894,818B
 ; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/03253
 ; FILING DATE: 07-NOV-1996

PRIOR APPLICATION DATA: JP 323285/1995
 ; APPLICATION NUMBER: 12-DEC-1995
 ; FILING DATE: 12-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: TAKAKURA-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1977 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; US-08-894-818B-6

Alignment Scores:

Pred. No.: 2,03e-30 Length: 1977
 Score: 401.50 Matches: 130
 Percent Similarity: 43.96% Conservative: 63
 Best Local Similarity: 29.61% Mismatches: 167
 Query Match: 17.74% Indels: 79
 DB: 3 Gaps: 17

US-09-985-689A-3 (1-433) x US-08-894-818B-6 (1-1977)
 QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaAlaAspThrGlyLeuAspThr 35
 Db 454 AACTGGGATATGATGGTTCTGAATCAATAGGAATTAATGACATGGAATTGAC--- 510
 QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 511 -----GCTTCTCATCCAGATCTCCAAGGAAGAATA----- 540
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 541 ATTGGGTGGGTAGATTGTTGCAATGGTAGGAGTTATCCATACGATGACCATGGACATGGA 600
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 601 ACTCATGTAGCTTCAATAGCAGCTGTACTGGAGCAGCAAGTAATGGCAAGTACAGGGA 660
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGGTTCTGGAAGC 720
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 721 ATATCTACTATAATTAAGGAGTTGAGTGGCGGTTGATACAAAGATTAAGTACGGAAT 780
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 781 AAGGTCATTAAATCTTCTCTGTTCAAGCCAGAGCTCCGACGGAACCGACTCCCTCAGT 840
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 841 CAGGCGGTCAACACCGCTGGGACCGCGGTAGTAGTCTGGTCCGCCCGGCAACAGC 900
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 901 GGGCCGAACACCTACACCGCTCGGCTCACCGCGCGGAGCAAGGTCTATAACCGTCGT 960
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
 Db 961 GCA-----GTTGACAGCAACGACACATCGCC 987
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 988 AGCTTCTCAGCAGGGGACCGACCGGGACGGAAGGCTCAGCGGAAGTCTCGTCCGCC 1047
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 1048 GCGTTGACATCATAGCCCGCGCGCCAGC-----GGAACCAAGCATGGGCCCGCC 1098
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1099 ATAAACGACTACTACACCAAGGCTCTGGACCAAGCATGCCCGCACGTTTCGGGC 1158
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 1159 GTTGGCGCGCTCATCTCCAG-----GCCACCGAGCTGGACCGCGGACCAAGGTG 1209
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
 Db 1210 AAGACCGCTCATCGAGACCGCGACATAGTCCCGCCCAAGAGATACGGACATCGCC 1269
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 1270 TAC-----GGTGGGTAGGTGACGTCATCAAGGCC-----ATCAAG 1308
 QY 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGln-----LysAlaThr 333
 Db 1309 TAGCAGACTACCCCAAGCTCACCCTTCACCGCTCGTCCCGCAGCAAGGAAGCGCCACC 1368
 QY 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
 Db 1369 CACACCTTCGACGTCAGCGCGCGCACCTTCGTCGTGACCGCCACCTCTACTGGGAC----- 1422

QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1423 -----ACGGCTCGAGCGACATCGACCTCTACCTCTACCGAC 1458

QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTTPAsp 393
 Db 1459 CCCAAC-----GGACAGAGTTGACTACTCTACCCGCTACTAC----- 1500

QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
 Db 1501 -----GGCTTCGAGAGGTCGCTACTACAAACCCGACCCCGAACTCGAGCGTC 1551

QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1552 AAGTCTCAGCTAC-----AAGGCGCGCGGAACTACCAAGTCTCAGCTGTC 1599

RESULT 12
 US-09-000-016-3
 ; Sequence 3, Application US/09000016
 ; Patent No. 6143541
 ; GENERAL INFORMATION:
 ; APPLICANT: Akira ARISAWA et al.
 ; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
 ; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
 ; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
 ; STREET: 2033 K Street, N.W., #800
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/000,016
 ; FILING DATE: January 30, 1998
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-721-8200
 ; TELEFAX: 202-721-8250
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2539 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Streptomyces viridosporus
 ; STRAIN: A-914
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 338...2539
 ; IDENTIFICATION METHOD: E
 US-09-000-016-3
 Alignment Scores: 5,94e-25 Length: 2539
 Pred. No.: 348.50 Matches: 125
 Score:

Percent Similarity: 38.99% Conservative: 45
 Best Local Similarity: 28.67% Mismatches: 149
 Query Match: 15.40% Indels: 117
 DB: 3 Gaps: 17

US-09-985-689A-3 (1-433) x US-09-000-016-3 (1-2539)

QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
 Db 998 GGCTACGACGCGCAAGGGCGTGAGATCGCGTCTGGACACCGGTGTCGACACGAGC--- 1054

QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
 Db 1055 -----CATCGGACCTGAAGGCGCGGTGACCGCTCCAAGAACTTCACC 1099

QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 1100 GCGCGCGCGCGCGCGGACAGGTGGCGCCACCGACCCCGCTCGCTCGATCGCGGCG 1159

QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
 Db 1160 GGCACGGCGCGCCAGTCCAGGGCAAGTACAAAGGGCGTCGACCGCGCGCGCATCTC 1219

QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
 Db 1220 AACGCGAAGTCTCTCGACGACTCC-----GGTTTCGCGCAGCATCCGCGCATC 1267

QY 113 LeuPheSerGlnAlaTTPAsnAla-----GlyAlaArgIleHisThrAsnSerTTPGly 130
 Db 1268 CTCGCGGCGCATGGAGTGGCGCGCGCGAGGGCGCGCGCGCTCGTCACCATGAGCTGGGC 1327

QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
 Db 1328 GGCATGGACACACCGGAGACCGACCCGCTGGAGCGCGCG----- 1366

QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
 Db 1367 GTCGACAGCTGTCCGCGCGAAGGGCGTCTGTCCTCCATCGCGCGCGCAACAGGGC 1426

QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
 Db 1427 CCGGAGTCCG-----ATCGGTTCGCGCGCAGCGCGCGCGCTCACCGTCGCGGCGC 1480

QY 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
 Db 1481 -----GTCGACGACAGGCAAGCTCGCGGAC 1507

QY 204 PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 1508 TTCTCTCTCACCGCGCGCGCTCGCGCGCGCGCATCAAGCGCGAGCTCACCGCTCCC 1567

QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTTPAlaAsn 242
 Db 1568 GCGGTGGACATCAGCGCGCTCGCGGAGGGCAACGACATCGCGCGGAGGTGGTGAG 1627

QY 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1628 GGACCGCGCGGCTACATGACCATCTCCGCGCGCGCGCATGATGGCGACCCCGCGCGGCG 1687

QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
 Db 1688 GCGCGCGCGCTCTGAGCAGCAG-----CACCCCGAC 1720

QY 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1721 TGGACCTCCGCGCAACTGAGGGCGGCTCACCGCTCCACCAAGGGCGGC---AAGTAC 1777

QY 300 ProSerGlyAspGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 Db 1778 ACCCGGTTTCGAGCAGGGTTCGGCGCGGATCCAGGCCGCGCAAGGGCGCTCCAG----- 1828

QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
 Db 1828


```

Db 1829 -----CAGACCGTG 1837
Qy 340 AlaGlyLysProLeuLysIleSerLeu-----ValTrp-----ThrAspAla 353
Db 1838 ATCGCGGACCGGCTCGGTGAGTTCGGCGTCCAGCAGTGGCGGCACACCGACGACGAG 1897
Qy 354 ProGlySerThrThrAlaSerThrLeuValAsnAspLeuValIle----- 371
Db 1898 CCGGTACACAGCAGCTACCTACCGAACCTTCGGCACCAGGACGTACCGCTGAAGCTG 1957
Qy 372 -----ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrPro 387
Db 1958 ACGTCGACCGCCACCGACCCCAAGGCAAG----- 1987
Qy 388 TyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln 407
Db 1988 -----GCGGCCCG 1996
Qy 408 SerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProSerGly 423
Db 1997 GCGGCTCTTCACGCTGGCGGCCACCGACCGGTGACCGTCCCGGGCGGC 2044

RESULT 13
; Sequence 3, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E

```

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-514-340-3
Alignment Scores:
Pred. No.: 5,94E-25 Length: 2539
Score: 348.50 Matches: 125
Percent Similarity: 38.99% Conservative: 45
Best Local Similarity: 28.67% Mismatches: 119
Query Match: 15.40% Indels: 117
DB: 4 Gaps: 17
US-09-985-689A-3 (1-433) x US-09-514-340-3 (1-2539)
Qy 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGTACGACGGCAGGCGGTGAAGATCGCGCTCCGACACCGGTGTCACACGAGC--- 1054
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTGAAGGCGCGGTGACCGGTCCCAAGAACTTCACC 1099
Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCGCGCGCGCGCGCGGACAAAGTGGGCGCACCGCACCGTCCCTCGCTCGCGGGG 1159
Qy 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCGCGCCAGTCCAAAGGCAAGTACAAGGGCGTCCACCGCGCGCGATCCTC 1219
Qy 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 1220 ACGGCAAGTCTCTGACGACTCC-----GGTTTCGGCGACACTCCGGGCATC 1267
Qy 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1268 CTGCGCGCATGAGTGGCGCGCGCGCGAGGCGCGACGTCTCACCATGAGCTCGGCGC 1327
Qy 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 1328 GGCATGACACACCGGAGACCGACCGCTGGAGGGCGGCG----- 1366
Qy 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 GTCGACAAAGCTGTCCGCGGAGAGGCGTCTGTTCGCCATCGCGCGCGGCAACGAGGCG 1426
Qy 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTCG-----ATCGGTTCCCGCGGACGCGCGCGCTCACCCTCGCGCGCGC 1480
Qy 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
Db 1481 -----GTCGACGACAGGACAAAGCTCGCGCGAC 1507
Qy 204 PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 1508 TTCTCTCCACCGCGCGCGCGCTCGCGGACGCGCGCATCAAGCGCGACGTCCCGCTCC 1567
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGACATCACCGCGCGCGCTCGCGGAGGCGGACACATCGCGCAGAGGTCGTGTAG 1627
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACCGCGCGGCTACATGACCATCTCCGGCACGTCTGATGCGACCGCGCGCGCGCGG 1687
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
Db 1688 GCGCGCGCGCTCTCTGAAGCAGCAG-----CACCCCGGAC 1720
Qy 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TGGACCTCCCGCGAACTGAAGCGCGCGCTCACCGGCTCCACCAAGGCGCGCG---AAGTAC 1777

```

```
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGCTTCAGCAGCGGCTCGGCGCGGATCCAGCGCGACAGCGCTCCAG-----1828
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1829 -----CAGACCGGTG 1837
QY 340 AlaGlyLysProLeuLysLeuSerLeu-----ValTrp-----ThrAspAla 353
Db 1838 ATCGCGCAGCCGCTCGGTGAGCTTCGGCGTCCAGAGTGGCGGCACCGACCGACGAG 1897
QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuValIle-----371
Db 1898 CCGGTACCAAGCAGCTGACCTACCGCAACCTCGGCACCCAGGACGTACGCTGAAGCTG 1957
QY 372 -----ThrAlaProLengGlnLysTyrValGlyAsnAspPheSerTyrPro 387
Db 1958 ACGTGACCGCCACCGACCCCAAGGCAAG-----1987
QY 388 TyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln 407
Db 1988 -----CGGCGCCCG 1996
QY 408 SerGlyThrTyrThrThrLeuValGlnAlaTyrAsnValProSerGly 423
Db 1997 GCGGGCTTCTTCAGCTGGCGCGCCACCGAGTGCACGCTCCCGCGGCGC 2044

RESULT 14
US-09-000-016-1
; Sequence 1, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIBINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
```

```
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; STRAIN:
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores:
Pred. No.: 2,19e-24 Length: 2809
Score: 343.50 Matches: 124
Percent Similarity: 38.76% Conservative: 45
Best Local Similarity: 28.44% Mismatches: 150
Query Match: 15.18% Indels: 117
DB: 3 Gaps: 17
US-09-985-689A-3 (1-433) x US-09-000-016-1 (1-2809)
QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGCAAGGCGGTGAAGATCGCTCTGGACACCGGTGTGACACGAGC---1054
QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTGAGGCGCGGTGACCGGTCCAAAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCCGCGCGCGCGCGCGGACAGGTGGCGCACCGCACCGCTCGCTCGATCGCGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCGCGCGCGGTCCAGGGCAAGTACAGGGCGTGCACCGCGCGCGGATCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 1220 AACGGCAAGGTCTCGACGACTCC-----GGTTTCGGCGGACGACTCGGCGATC 1267
QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1268 CTCGGCGCATGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 1328 GGCATGGACACACCGGAGACCGACCGCTGGAGGCGGCG-----1366
QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 GTCGCAAGCTGTCCGCCGAGAGGGGTCTGTGTCCCATCGCGCGCGCGCAAGGGGC 1426
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTGC-----ATCGGTTCGGCGGACGCGCGCGCGCGCGCGCGCGCG 1480
QY 184 ThrGluAsnTyrArgProSerPheIleAlaAspAsnProAsnHisIleAlaGln 203
Db 1481 -----GTCGACGACAGGACCAAGCTCGCGAC 1507
QY 204 PheSerSerArgGly-----AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 1508 TTCTCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCC 1567
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGGACATCACGCGCGCGCTCGCGCGAGGCGCAACGACATCGGCGCAGAGGTGCGTGA 1627
```

QY 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACGGCGCGGTACATGACCATCTCCGGCAGCTGATGGCAGCCCGCAGCTCGCGGC 1687
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyLeuThrProLysProSer 282
Db 1688 GCGCGCGCCTCTGAGCAGCAG-----CACCCCGAC 1720
QY 283 LeuIleLysAlaAlaLeu-----lleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TGGACCTCCGCGAACTGAAGCGCGCTCACCGCTCCACCAAGCGCGC---AAGTAC 1777
QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGCTTCGACAGGGTTCGGCGCGATCCAGCGCAGCAGCGCTCAG----- 1828
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1829 -----CAGACCGTG 1837
QY 340 AlaGlyLysProLeuLysIleSerLeu-----ValTyr-----ThrAspAla 353
Db 1838 ATCGCGCAGCGGTCTCGGTGAGCTTCGGCGTCCAGAGTGGCCGACCGCAGCAGCAG 1897
QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle----- 371
Db 1898 CCGGTCAACAGCAGCTGACCTACCGCACTCCGACCCGACCGCAGCTCAGCTGAAGCTG 1957
QY 372 -----ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrPro 387
Db 1958 ACCTGACCGCCACCGACCCCAAGGCAAG----- 1987
QY 388 TyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln 407
Db 1988 -----CGCGCCCG 1996
QY 408 SerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProSerGly 423
Db 1997 GCGGGCTTCTTCAGCTGGCGCCACCGCAGCTGACCGTCCCGCGCGGC 2044

RESULT 15
US-09-514-340-1
Sequence 1, Application US/09514340
Patent No. 6361987
GENERAL INFORMATION:
APPLICANT: Akira ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wengderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN: <Unknown>
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores:
Pred. No.: 2,19e-24 Length: 2809
Score: 343.50 Matches: 124
Percent Similarity: 38.76% Conservative: 45
Best Local Similarity: 28.44% Mismatches: 150
Query Match: 15.18% Indels: 117
Gaps: 17
US-09-985-689A-3 (1-433) x US-09-514-340-1 (1-2809)
QY 18 GlyLeuTyrGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGCAAGGCGGTGAAGATCGCGCTCTGGACACCGCGTGTGCACGAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTGAAGGGCGGTGACCGCTCCAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCGCGCGCGCGCGCGCGGACAGGTGGCGCAGCGCACCGCACCTCGCTCGATCGCGGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCGCGCGCGCGCGCAAGTACAGGGCGTGCACCGCGCGCGGATCCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db 1220 AACGGCAAGGTCTCTCGACGACTCC-----GGTTTCGGCAGCAGCTCCGCGCATC 1267
QY 113 LeuPheSerGlnAlaTyrAsnAla-----GlyAlaArgIleHisThrAsnSerTyrGly 130
Db 1268 CTGCGCGCATGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 1328 GGCATGGACACACCGGAGACCGCGCTGGAGCGCGG----- 1366
QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 GTGACAAGCTGTCCGCGGAGAGGGGCTCTTTCGCCATCGCGCGCGCGCGCGCGCGCG 1426
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTCTG-----ATCGGTTCCCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCG 1480
QY 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203

```

Db 1481 -----GTCGACGACAGGACAAAGCTGCCGAC 1507
Qy 204 PheSerSerArggly---AlaThrArgAspGlyArgIleLeuSerProaspValThrAlaPro 222
Db 1508 TTTCTCTCCACCGCCCGCTCGCGAGCGGGCCATCAAGCCGACGTACCGCTCCC 1567
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGGACATCACGCCCGCTCGCGGAGGGCAACGACATCGGCCAGAGGTCCGTGAG 1627
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGCACCGCGCGGTACATGACCATCTCCGGCACGTGATGCGGCCCGCACGTCCGGGCGC 1687
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
Db 1688 GCGCGCGCCCTCTGAAGCAGCAG-----CACCCCGAC 1720
Qy 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TGGACCTCCGCCAACTGAAGGGCGGCTCACCGGTCCACCAAGGGCGGC---AAGTAC 1777
Qy 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGGTTCGAGCAGGGTTCCGGCGCGATCCAGCGCGACAAAGCGCTCCAG----- 1828
Qy 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1829 -----CAGACCGTG 1837
Qy 340 AlaGlyLysProLeuLysIleSerLeu-----ValTrp-----ThrAspAla 353
Db 1838 ATCGCCGACCGGTCCTCGGTGAGTTCGGCGTCCAGCAGTGGCGGCACACCGACGACGAG 1897
Qy 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle----- 371
Db 1898 CCGGTCCACCAAGAGCTGACTACCGCACTCGGCACCCAGGACGTACGCTGAAGCTG 1957
Qy 372 -----ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrPro 387
Db 1958 ACGTCGACCGGCACCGACCCCAAGGGCAAG----- 1987
Qy 388 TyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln 407
Db 1988 -----GCGGCCCG 1996
Qy 408 SerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProSerGly 423
Db 1997 GCGGGCTTCTTCACGCTGGGCGGCCACCGCGTGACCGTCCCGCGGGGC 2044

```

Search completed: April 4, 2004, 12:02:03
Job time : 90.4512 secs

Sequence 1, Appli
Sequence 11, Appli
Sequence 2, Appli
Sequence 15, Appli
Sequence 1, Appli
Sequence 1934, Ap
Sequence 5701, Ap
Sequence 3306, Ap
Sequence 1, Appli
Sequence 29, Appli
Sequence 133, App
Sequence 2, Appli
Sequence 5384, Ap
Sequence 17, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 6, Appli
Sequence 20, Appli
Sequence 20, Appli
Sequence 33, Appli
Sequence 12, Appli
Sequence 1, Appli
Sequence 5, Appli
Sequence 18, Appli
Sequence 18, Appli
Sequence 1938, Ap
Sequence 1938, Ap
Sequence 112429,
Sequence 34, Appli
Sequence 30, Appli
Sequence 9, Appli
Sequence 13, Appli
Sequence 5689, Ap
Sequence 3, Appli
Sequence 953, App
Sequence 19, Appli
Sequence 19, Appli
Sequence 5, Appli
Sequence 98, Appli

1 1994.5 88.1 1305 15 US-10-385-662-1
2 443 19.6 1977 13 US-10-090-624-11
3 422.5 18.7 1236 13 US-10-090-624-2
4 422.5 18.7 1962 13 US-10-090-624-15
5 346.5 15.3 135638 14 US-10-314-857-1
6 336 14.8 1329 9 US-09-974-300-1334
7 329 14.5 9025608 14 US-10-156-761-5701
8 329 14.5 9025608 14 US-10-156-761-1
9 327 14.4 3417 14 US-10-156-761-3306
10 337 14.4 9025608 14 US-10-156-761-1
11 305 13.5 4765 13 US-10-090-624-5
12 296.5 13.1 3743 13 US-09-927-827-29
13 295 13.1 1560 15 US-10-084-846A-113
14 295 13.0 59816 15 US-10-084-846A-2
15 295 13.0 59816 15 US-10-084-846A-2
16 282.5 12.5 3303 14 US-10-156-761-5384
17 281.5 12.4 2166 12 US-10-344-231-17
18 281.5 12.4 2166 12 US-10-363-332A-17
19 269 11.9 1140 8 US-08-322-678-11
20 269 11.9 1140 16 US-10-323-324-11
21 269 11.9 1143 14 US-10-313-853-6
22 269 11.9 2588 12 US-10-344-231-20
23 269 11.9 2588 12 US-10-363-332A-20
24 269 11.9 3788 10 US-09-927-827-33
25 268 11.8 1140 8 US-08-322-678-12
26 268 11.8 1140 16 US-10-323-324-12
27 260.5 11.5 1506 9 US-09-966-921A-1
28 260.5 11.5 1330 9 US-09-966-921A-5
29 259.5 11.5 840 14 US-10-209-812-1
30 258.5 11.4 2267 12 US-10-344-231-18
31 258.5 11.4 2267 12 US-10-363-332A-18
32 252.5 11.2 1971 9 US-09-974-300-1935
33 252 11.1 1485 9 US-09-974-300-1938
34 247.5 10.9 2192 12 US-10-424-559-112429
35 247 10.7 3884 10 US-09-927-827-34
36 242.5 10.7 3452 10 US-09-927-827-30
37 242 10.7 1140 15 US-10-146-905A-9
38 238 10.5 1140 14 US-09-920-118-13
39 232 10.3 1332 14 US-10-156-761-5689
40 231.5 10.2 4338 12 US-09-891-711-3
41 231.5 10.2 4338 12 US-10-342-887-953
42 230 10.2 1575 12 US-10-344-231-19
43 230 10.2 1575 12 US-10-363-332A-19
44 228 10.1 4198 9 US-09-891-711-5
45 227.5 10.1 522 10 US-09-824-893A-98

ALIGNMENTS

US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US2004002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938US0
; CURRENT APPLICATION NUMBER: US/10/385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

RESULT 1

US-10-385-662-1

; Sequence 1, Application US/10385662
; Publication No. US2004002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938US0
; CURRENT APPLICATION NUMBER: US/10/385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 327.248 Seconds
(without alignments)
4948.852 Million cell updates/sec

Title: US-09-985-689A-3
Perfect score: 2263
Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVPSGPQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO/spool/US09985689/runat_31032004_161809_4271/app_query.fasta_1.3498
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINWATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -SPART=1 -END=-1 -MATRIX=blsum62
-TRANS=human40 cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODS=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09985689@cgn_1_1_601@runat_31032004_161809_4271
-ICPU=3 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result %
No. Score Match Length DB ID Description

PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Bacillus sp. KSM-KP43
; FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1305)
OTHER INFORMATION:

US-10-385-662-1

Alignment Scores:
Pred. No.: Length: 1305
Score: 1994.50 Matches: 380
Percent Similarity: 94.01% Conservativity: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.14% Indels: 1
DB: Gaps: 15

US-09-985-689A-3 (1-433) x US-10-385-662-1 (1-1305)

QY 1 AsnAspValAlaArgGlyLeuVallayalaaspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AATGATTGGCGGTGAATTCCTCAAGCGGATGGTGCTCAGACGACTACGGTTGTAT 60
QY 21 GlyGlnGlyValValAlalvalalpThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAGGCAGATCATCCGTTGCCGATACCAGGCTTGATACAGSTCGCAAATGACAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrrAlaLeuGlyArgThrAsn 60
Db 121 TCATGTCATGAAGCCITTCGCGGAAATAATTAATGATATATGATTCGGACGACGAAT 180
QY 61 AsnAlaAspProAsnGlyHisGlyThrHisValAlaGlyservalLeuGlyAsn--- 79
Db 181 AATGCCAATGATACGAATGGTCATGTCATGTCATGTCGCTGGCTCCGTATTAGGAACGGC 240
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerileMetAspSer 99
Db 241 TCCACTAATAAGAATGGCGCTTAGGGCAATCTAGICTTCCAATCTATCATGGATAGC 300
QY 100 SerGlyGlyLeuGlyLeuProserAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 301 GTGGGGGACTTGGAGGACTACCTTCGAATCTGCAAACCTTATTGAGCAAGCACATACAGT 360
QY 120 AlaglyAlaArgIleHistrAnsenSertPgylAlaProValasnGlyValatryrAla 139
Db 361 GCRTGGCCAGAATTATCAAAAACCTCTGGGAGCAGCAGTGAATGGGCTTACCAACA 420
QY 140 AnSenArGlnValaspGlutryrValargAsnasnAspMetThrValuPheAlala 159
Db 421 GAITCCAGAAATGTGATGACTATGTGCGCAAAATGATATGACGATCCTTTTCGCTGCC 480
QY 160 GlyasnglugluproasenSerglythrileserAlaprogllyThrAlalyshsnlaile 179
Db 481 GGGAAATGAAGCCGAAACCGCGGAACCATCAGTGCACCGAGGACAGCTAAAAATGAATA 540
QY 180 ThrValglyAlatrGluAsntyrrArgproferPheGlySerifleAlahspAsnpRoasn 199
Db 541 ACATCGAGCTACGGAAACCTTCGCCCCAAGCTTTGGGTCTATGCGGACATATCAAC 600
QY 200 HistleAlaGlnPheSerSerArgGlyAlathArgaspGlyArgyleysroaspval 219
Db 601 CATGTGGCAGATTTCTCTTCAGCTGGACCGACAAAGATGGACGATCAAAACGGATGTC 660
QY 220 ThralaproglythrPhelileUserAlargserserleualapropasperPhe 239
Db 661 ATGGACCGGAAGCTTCATATCATCATGCAAGATCTCTTCCTGACCGGATTCCTTC 720
QY 240 TrpAlaAsntyrrAsnsLystrValatyrmetyrglytyrservmetlatrProfile 259

721 TGCGGAACCATGACAGTAATATGCTCATATGGTGGACCGTCCATGGCTCACCGATC 780

260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleylsanargGlylleThrPro 279

781 GTTGTGAAACAGTGGCACACAGTTCGTGAGCATTTTTTGAAAAACAGAGCATCACACCA 840

280 LysProSerleulleLysAlaaleuleAlaGlyAlaThrAspValGlyLeuGlyTYR 299

841 AAGCCTTCTATTAAAGCGGCACTGATTGCCGGTGCAGCTGACATCGGCCTTGCTAC 900

300 ProSerGlyAspGlnGlyTtpGlyArgValThrLeuaspLysSerleuAsnValAlaTYR 319

901 CGRAACGGTAACCAAGGATGGGACGAGTGACATTTGGATAAATCCCCTGAACGTTGCCAT 960

320 VallaenGluAlaThrAlaLeuThrTheGlyGlnLysAlaThrTVrSerPheGlnThrGln 339

961 GFRAACGAGTCCAGTTCTCTATCCACGACGCAAAAACGACGACTACTCGTTTACTGCTACT 1020

340 AlaglyLysProLeuLysIleserLeuValtrpThrAspalaproGlyserThrAla 359

1021 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTCGGACGACAACTGCT 1080

360 SerTYrThrLeuValasnAspleuasplauvulllethrAlaProasnGlyGlnLYseTYR 379

1081 TCCGTPAACCGTGTCTCAATGATCTGGACCTTCTCATATCCGCTCCAAATGGCACACATAT 1140

380 ValGlyAsnAspPheSerTYrProtzyrAspnAsnPsrAspGlyArgAsnAsnValGlu 399

1141 GTAGGAATGACTTTACTTCGCCATACATGATACTGGGATGCCGCAATAAGCTAGAA 1200

400 AsnValpheilleAsnalaproGlnserGlyThrTYrThrIleGluValGlnAlaTYrasn 419

1201 AATGTATTTTATTAATGCAACCAAGCGGAGCGTATACAAATTGAGGTACAGGCTTATAAC 1260

420 ValProSerGlyProGlnArgPheSerleuAlaileValHis 433

1261 GTACCGGTGGACCAACAGACCTTCTCGTTGGCAATTGTGAAT 1302

RESULT 2
US-10-090-624-11
Sequence 11, Application US/10090624
Publication No. US2002013235A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPlicant: MORISHITA, Mio
Applicant: SHIMOJO, Tomoko
Applicant: ASADA, Kiyozo
Applicant: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA=6
CURRENT APPLICATION NUMBER: US/10/090, 624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445, 472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 11
LENGTH: 1977
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: Length: 2,82e+38
Score: 443.00 Matches: 137
Percent Similarity: 46.03% Conservativity: 66
Best Local Similarity: 31.07% Mismatches: 170
Query Match: 19.58% Indels: 68
DB: Caps: 16

US-09-985-689A-3 (1-433) x US-10-090-624-11 (1-1977)

Qy 8 ValysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyGlnValVala 27
Db 433 ATAGGGCCGATACCGTCTGGAATCTCTCGGTACAGCAAGCGGTGTGGTGGCC 492
Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
Db 493 ATCTCGATACGGGTATAGACGGAAC-----CACCCCGATCTGAAG 534
Qy 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
Db 535 GGCAGGTCATAGGCTGGTACGCGCTCAACGGCAGGTGACCCCTCAGATGACGAG 594
Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
Db 595 GGACACGGAACCCACGTTGGGGTATGTTCCCGGAACCCGCGCAGCGTTAACTCCAGTAC 654
Qy 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
Db 655 ATAGCGGTGCGCCCGCGCGGAAGCTCGTCGGCTCAAGGTTCTCGGTGCGGACGGTTCG 714
Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 715 GGAACGCTCTCCACCATCATCGGGGTGTTGACTGGTGGTCCAGAACAGGACAGTAC 774
Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 775 GGGATAGGGTTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGGGAACCGACTCC 834
Qy 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
Db 835 CTCAGTCAGCGCTCAACAGCGCTGGACCGCGGTATAGTAGTCTGGTCCGCGCGGC 894
Qy 161 ArgGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 895 AACAGCGGGCGGAACACCATACCGTCCGCTCACCGCGCGCGGAGCAAGTCTATAACC 954
Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 955 GTCGGTGCA-----GTTGACGCAACGCAAC 981
Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 982 ATCGCCAGTCTTCAGCAGGGGACCGACCGGAGGAGGCTCAAGCCGGAAGTCGTC 1041
Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 1042 GCCCCCGCGGTTGACATCATGCCCGCGCGCGCAGC-----GGAAACGAGCATGGGC 1092
Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 1093 ACCCCGATAAAGCTACTACACCAAGGCTCTGGAACCAAGCATGGCCACCCCGCAGCTT 1152
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
Db 1153 TCGGCGGTGGCGCGCTCATCTCCAG-----GCCACCCGAGCTGACCCCGGAC 1203
Qy 280 -----LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
Db 1204 AAGGTGAAGACCGCCTCATGAGACCGCGGACATAGTCGCCCCCAAGGATAGCGGAC 1263
Qy 297 LeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsn 316
Db 1264 ATCGCTAC-----GGTGGGTAGGGTGAAGCTCTACAGGCC----- 1302
Qy 317 ValAlaTyrValAsnGluAlaThrAlaLeuThrThrGlyGln-----Lys 331
Db 1303 ATCAAGTACAGCAGTACGCCAAGCTCACCTTACCGGCTCGCTCGCGCAACAGGGAAGC 1362
Qy 332 AlaThrTyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTrpThr 351
Db 1362

US-09-985-689A-3 (1-433) x US-10-090-624-2 (1-1236)

Qy 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTTGGATATGATGTTCTGGATCAACATAGGAATAATTGACACTGGAATTGAC--- 114
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAla 55
Db 115 -----GCTTCTCATCCAGATCTCCCAAGGAAGAAGTA----- 144
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATTGGGTGGTAGATTGTTGCAATGTAGGAGTATCCATGATGACCATGACATGGA 204
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 1452

Db 1363 GCCACCCACACCTTCGAGCTGAGCGGCCACCTTCTGTGACGGCCACCTCTACTGGGAC 1422
Qy 352 AspAlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValle 371
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTC 1452
Qy 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
Db 1453 TAGCACCCCAAC-----GGGAACGAGGTGACTCTCTTACACCCGCTTAC 1497
Qy 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
Db 1498 TAC-----GGCTTCGAGAGGTGCGCTACTACACCCGCGCGGACCTGG 1545
Qy 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
Db 1546 ACGGTCAAGTCTGTCAGCTAC-----AAGGCGCGCGGCACTACGAGTGCAGCTC 1596
Qy 432 Val 432
Db 1597 GTC 1599

RESULT 3
US-10-090-624-2
; Sequence 2, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090, 624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2

Alignment Scores:
Pred. No.: 2,53e-36 Length: 1236
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.67% Indels: 79
DB: 13 Gaps: 17


```
Db 901 GGACCTAACAGTATACATCGTTCTCCAGCAGCTGCAAGCAAAAGTTATACAGTTGGA 960
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 961 GCCGTTGACAGTAT-----GATGTTATAACA 987
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 988 AGCTTCTCAACAGAGGCCAACTGCACAGCGCAGGCTTAAGCCTGAGGTGTGTGCTCCA 1047
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyrAlaAsn 242
Db 1048 GGAACCTGGATTAATGTCGCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 1098
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1099 ATTAATGACTATTACAGCAGCTCTCTGGGACATCAATGGCACTCCTCAGTAGCTGGT 1158
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 1159 ATTGCAGCCCTCTTGCTCCAA-----GCACACCGAGCTGGATCCAGCAAAAGTA 1209
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 1210 AAAACAGCCCTCATAGAAACTGCTGATATCGTAAAGCCAGATGAATAGCGGATATAGCC 1269
Qy 299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 1270 TAC-----GGTGCAGGTAGGTTTAATGCATACAAAGGTCATAAACC----- 1308
Qy 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
Db 1309 TACGATAACTATGCAAGCTAGTGTCTACTGGATATGTTGCCAACAAAGGCGACCAAACT 1368
Qy 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThrAspAla 353
Db 1369 CACCAAGTTCGTTATTAGCGGAGCTTCGTTCTGTAACCTGCACATTATATCTGGCAATGCC 1428
Qy 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 1429 AAT-----AGCGACCTTGATCTTTACCTCTACGAT 1458
Qy 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnThrAsp 393
Db 1459 CCCAATGGAACCAACCGGTT-----GACTACTCTTAC-----ACC 1491
Qy 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
Db 1492 GCCTACTATGATTCGAAAGGTTGGTTATTACAAACCACTGATGGNACATGGACAAATT 1551
Qy 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGGTTGTAAGTAC-----AGCGGAAGTGCAAACTATCAAGTAGATGTGGTA 1599
```

RESULT 5

```
US-10-314-657-1
; Sequence 1, Application US/10314657
; Publication No. US20030175888A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: TANG, Gong-Li
; APPLICANT: CHENG, Yi-Qiang
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; TITLE OF INVENTION: Syntheses and Methods of Use
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1
```

```
Alignment Scores:
Pred. No.: 4,928-25 Length: 135638
Score: 346.50 Matches: 140
Percent Similarity: 42.39% Conservative: 55
Best Local Similarity: 30.43% Mismatches: 154
Query Match: 15.31% Indels: 111
DB: 14 Gaps: 23
```

US-09-985-689A-3 (1-433) x US-10-314-657-1 (1-135638)

```
Qy 6 GlyIleValLysAlaAspValAlaGlnAsnAsn----- 16
Db 8067 GGCAGGGTGAAGGCGATCTGGCCGACTCCACCGCCAGATCGGGCGCGCAGAGGTATGG 8126
Qy 17 ---TyrGlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 8127 GCGGAGGGCGACACCGGCGCAGGACGTGAAGGTGCGGATGCTGCACAGCGCGCGACACC 8186
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
Db 8187 -----GAACACCGGACCTGTGCGGCGAGGTGTCCGACAGCGCCAGC 8228
Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnProAsnGlyHisGlyThrHisValAla 73
Db 8229 TTCGTCCCGGCGGAGCAGCATCGCC---GACTACAAACGGCCAGCGCAGCATCGCC 8285
Qy 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88
Db 8286 TCGACCATCTGCGGACAGCGGCGCGCTCCGACGCAAGGAGCGGGGTGCGCTCCGCG 8345
Qy 89 AlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSer 108
Db 8346 GCGCGCTCTCGCTCGGCAAGGTGCTCACTCCGAGGCG-----AGCGCGCAGGAATCG 8399
Qy 109 AsnLeuAsnThrThrPheSerGlnAlaTyrAsnAla-----GlyAlaArgIleHis 125
Db 8400 TGGATC-----ATCGCGGCGATGGAGTGGGCGCGCCGCGCAGACGACGAGATCATC 8453
Qy 126 ThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsn-----SerArg 142
Db 8454 AGCATGAGCTGGCG-----GGCGCGGTGTACACAGACGACCGGATGAGCCAG 8501
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 8502 GCGTCGACGAACTACGCCACGACAGCGGCGGTGTGTCGTGATCGCGCGGCGACCGC 8561
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 8562 GCGCGCACTCC-----ATCAGACGCGCGGTGCGGACAGCTCCGCGGTGACCGTCGCG 8615
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 8616 GCC-----GTGCACTCCACCGACGCTCGCC 8642
Qy 203 GlnPheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAla 221
Db 8643 GACTTCTCAGCGAGGCGCGCGTGCAGCGGCGGCGGCTGAAGCGGAGATCACCGCG 8702
Qy 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAla 241
Db 8703 CCGCGGCTCGACATCGTCGCGGCGGCTCGCTCAAGCGCGGCTCCGGCTAC----- 8756
Qy 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 261
Db 8757 -----TACACCGATGAGCGGCACTCGATGGCGGACGCGCGCATCGCC 8801
```



```

; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701

Alignment Scores:
Pred. No.: 2,158-25 Length: 3624
Score: 329.00 Matches: 119
Percent Similarity: 42.75% Conservative: 49
Best Local Similarity: 30.28% Mismatches: 161
Query Match: 14.54% Indels: 64
DB: 14 Gaps: 16

US-09-985-689A-3 (1-433) x US-10-156-761-5701 (1-3624)
QY 8 ValLysAlaAspValAlaGlnAsn-----Tyr 17
DB 556 GTCGAGGCGGACATCGCGGAGGACGACGCGGATCGGTACGCGGCGCGCGCGGTGGAGCGCC 615
QY 18 GlyLeuTyrGlyGlnGlyClnValValAlaValAlaLeuThrGlyLeuAspThrGlyArg 37
DB 616 GGGCTCACGGGCGGCGGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCG 669
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
DB 670 -----ACTCACCCGACCTCGCGCGCGGTGTCGCGGAGCAAGACTTCATC 717
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
DB 718 GACGGGAGGAGGTCCGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGAC 777
QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
DB 778 GCGAGCGGCGCGCTCCGACGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 837
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
DB 838 GTCGCGCAAGTCTCAGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGAC 885
QY 113 LeuPheSerGlnAlaThrAsnAla-----GlyAlaArgIleHisThrAsnSerTyr 129
DB 886 ATCGGGGATGATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 945
QY 130 GlyAla---ProValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrVal 148
DB 946 GGATCGACCGGCGGACCGGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1005
QY 149 ArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThr 168
DB 1006 GAGGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1062
QY 169 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArg 188
DB 1063 ATCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1107
QY 189 ProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGly 208

```

RESULT 8

```

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1,798-20 Length: 9025608
Score: 329.00 Matches: 119

```

Percent Similarity:	42.75%	Conservative:	49
Best Local Similarity:	30.28%	Mismatches:	161
Query Match:	14.54%	Indels:	64
DB:	14	Gaps:	16
US-09-985-689A-3 (1-433) x US-10-156-761-1 (1-9025608)			
QY	8	ValLysAlaAspValAlaGlnAsn	-----Tyr 17
DB	6918813	GTGAGGCGGACATGCGCGAGAGCAACGCGCAGATCGGTACGCGGCGCGCGGTGGAGCGCC	6918872
QY	18	GlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg	37
DB	6918873	GGGCTCAGCGGCGACGCGCTCACCGTCGCGTCTCGACACCGCGGTGACAC	6918926
QY	38	AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly	57
DB	6918927	-----ACTCACCGGACCTCGCGCGCGCGGTCTCGCGAGCAAGCTTCATC	6918974
QY	58	ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu	77
DB	6918975	GACGGGAGGAGTCCCGGACCCGACCGGACCGGACCGGACCGTACCTCGACCGTCCGC	6919034
QY	78	Gly	-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnVal 92
DB	6919035	GGCAGCGCGCGCGCTCCGACGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	6919094
QY	93	PheGlnSerIleMetAspSerGlyGlyLeuGlyLysLeuProSerAsnLeuAsnThr	112
DB	6919095	GTGCGGMAAGTGTACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	6919142
QY	113	LeuPheSerGlnAlaThrAsnAla	-----GlyAlaArgIleHisThrAsnSerTrp 129
DB	6919143	ATCGCGGCGATGGAATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	6919202
QY	130	GlyAla	-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrVal 148
DB	6919203	GGATCGACGAGGCGGACGCGGACCGGACCGGACCGGACCGGACCGGACCGGAC	6919262
QY	149	ArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThr	168
DB	6919263	GAGGAGACCGCGCGCGCTTCGTCGTCGCGCGGGAACACCGGTGCCCGCTCC	6919319
QY	169	IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArg	188
DB	6919320	ATCGGCTCG	6919364
QY	189	ProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGly	208
DB	6919365	-----GTCGACTCATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	6919406
QY	209	AlaThrArg	-----AspGlyArgIleLysProAspValThrAlaProGlyThrPheIleLeu 227
DB	6919407	CCGCGCCACGCGGACAGCGCGCTCAAGCGCGACCTCCCGCGCGCGCGCGCGCG	6919466
QY	228	SerAlaArgSerSerLeuAlaProAspSerSerThrAlaAsnTyrAsnSerLysTyr	247
DB	6919467	CGCGCGCGCGCTCCAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	6919505
QY	248	AlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu	267
DB	6919506	ACCTCCATGAGCGGTACGTGATGCGACGCGCGCGCGCGCGCGCGCGCGCGCG	6919565
QY	268	ArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuIleLysAlaAla	287
DB	6919566	GCCGAGACGACCCCGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	6919625
QY	288	-----LeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGly	305
DB	6919626	GAGCAACTCGACCGCTCGTATATCATGTTGGGGCGGTCGGGTCAAGTGTGCCGCGCG	6919685
QY	306	TrpGly	-----ArgValThrLeuAspLysSerLeuAsnValAlaTyr-----ValAsn 321
DB	6919685	-----GAGCAACTCGACCGCTCGTATATCATGTTGGGGCGGTCGGGTCAAGTGTGCCGCGCG	6919745
QY	322	GluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAlaGly	341
DB	6919746	GACGCGGATGACCGCGTCCAG	6919796
QY	342	LysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSerTyr	361
DB	6919797	ACGACGCTGAGTGTAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	6919847
QY	362	ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 374	
DB	6919848	ACCTCG	6919877
RESULT 9			
US-10-156-761-3306			
; Sequence 3306, Application US/10156761			
; Publication No. US20030119018A1			
; GENERAL INFORMATION:			
; APPLICANT: OMURA, SATOSHI			
; APPLICANT: IKEDA, HARUO			
; APPLICANT: ISHIKAWA, JUN			
; APPLICANT: HOSIKAWA, HIROSHI			
; APPLICANT: SHIBA, TADAYOSHI			
; APPLICANT: SAKAKI, YOSHIYUKI			
; APPLICANT: HATTORI, MASAHIRA			
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES			
; FILE REFERENCE: 249-262			
; CURRENT APPLICATION NUMBER: US/10/156,761			
; CURRENT FILING DATE: 2002-05-29			
; PRIOR APPLICATION NUMBER: JP 2001-204089			
; PRIOR FILING DATE: 2001-05-30			
; PRIOR APPLICATION NUMBER: JP 2001-272697			
; PRIOR FILING DATE: 2001-08-02			
; NUMBER OF SEQ ID NOS: 15109			
; SEQ ID NO 3306			
; LENGTH: 3417			
; TYPE: DNA			
; ORGANISM: Streptomyces avermitilis			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (1)..(3417)			
US-10-156-761-3306			
Alignment Scores:			
Pred. No.:	3,276-25	Length:	3417
Score:	327.00	Matches:	114
Percent Similarity:	45.27%	Conservative:	44
Best Local Similarity:	32.66%	Mismatches:	126
Query Match:	14.45%	Indels:	67
DB:	14	Gaps:	12
US-09-985-689A-3 (1-433) x US-10-156-761-3306 (1-3417)			
QY	18	GlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg	37
DB	745	GGGTACGACGCGACGCGGCTCAAGATCGCGCTCTGGACACCGGTGTGAC	795
QY	38	AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly	57
DB	796	-----GCGACCCACCGGACCTCAGGACGAGGTGGCGGAGTCCAGAACTTCTCC	846
QY	58	ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu	77
DB	847	GCG	906
QY	78	Gly	-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
DB	907	GGACCG	966
QY	93	PheGlnSerIleMetAspSerGly	-----GlyLeu 103
DB	967	AACGGCAAGGTCTTCGACGACACCGCGCTCCGCGGACGACTCCGCGATCTCTCGCGCGCATG	1026

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 1

; LENGTH: 9025608

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

; US-10-156-761-1

Alignment Scores:

Pred. No.: 2,96e-20 Length: 9025608

Score: 327.00 Matches: 114

Percent Similarity: 45.2% Conservative: 44

Best Local Similarity: 32.6% Mismatches: 126

Query Match: 14.4% Indels: 67

DB: 14 Gaps: 12

US-09-985-689A-3 (1-433) X US-10-156-761-1 (1-9025608)

```
QY 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 1027 GAGTGGCGCGCGAGCAGCGCGCGCGCTCAACCTGAGCTTGGCGCGCGCAGCAC 1086
QY 123 gileHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSerArgG1 143
Db 1087 CCGAGATCGACCGCGTGAA-GCGAGGTCAACAGCTC----- 1125
QY 143 nValAspGluTrpValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 1126 -----TCCGAGGAGAGAGGCGATCTCTTCGCGATCGCGCGCGCAA 1166
QY 161 nGluGlyPro---AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
Db 1167 CGAAGGCGAGTTCGGCGAGCAGACATCGCTCCCGCGCAGCGCGCGCGCGCTCAC 1226
QY 180 rValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHi 200
Db 1227 CGTCGGCGCC-----GTGAACGACGCGCAA 1253
QY 200 sileAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220
Db 1254 GCTGGCGCTCTTCCAGCCCGCGCGCGCGCGCTCGACCGGCCATCAAGCCGACGTCAC 1313
QY 220 rAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTr 240
Db 1314 CGCACCGCGCGTGACATCACCGCGCC-----GCGCGCGCGCGCGCGCATCGA 1364
QY 240 pAlaAsnTyArgSerLys-----TyAlaTyMetGlyGlyThrSerMetAlaTh 257
Db 1365 CCAGGAGTCCGCGCAGAAAGCGGCGCTACCTCACCATCTCCGCTACGTCGATGGCGAC 1424
QY 257 rProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyI1 277
Db 1425 CCGCATGTCCGGCGCGCGCGCGCGCTCTCAAGCAGCAGCAC----- 1467
QY 277 eThrProLysProSerLeu-----IleLysAlaAlaLeuIleAlaGlyAlaThrAspVa 295
Db 1468 -----CCCAACTGGTCTGTCGCGAGCTCAAGCGCGCGCTGACCGCGTCCGCGAAG----- 1518
QY 295 lGlyLeuGlyTyProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLe 315
Db 1519 -GCGCGGAGTACACCGCTTCCAGCAGGCTCGGCGCGCTATCCGCTCGCAAGCGAT 1577
QY 315 uAsnValAlaTyValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTySe 335
Db 1578 CAAGCAGTCCGTGATCGCCCAACCGCACTCGGTGAGCTTCGGCAT-CCAGCAGTGGCGCG 1636
QY 335 rPheGlnThrGlnAlaGlyLysPro 343
Db 1637 ACACGACGACAAAGCGGTCAACCA 1661
```

RESULT 10

US-10-156-761-1/c

; Sequence 1, Application US/10156761

; Publication NO. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

```
QY 18 GlyLeuTyGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 4132505 GGGTACGACGCGCAAGCGCTCAAGATCGCGCTCTGGACACCGCGTGCAC----- 4132455
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGly 57
Db 4132454 -----GCGACCCACCGCGACCTCAAGGACCAAGTGGCGCGCTCTCTCC 4132404
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 4132403 GCGCGCGCGCGCGCGCGCGCGCTTCGGTTCAGCGCACGCGACGTCGCTTCATCGCGCG 4132344
QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 4132343 GGCACCGCGCGCAAGTCCAAACGCAAGTACAAAGGTGTGCGCGCGCGCGCATCTC 4132284
QY 93 PheGlnSerIleMetAspSerSerGly-----GlyLeu 103
Db 4132283 AACGGCAAGTCTCTGACGACACCGCGCTCCGCGCAGCATCGCGCATCTGCGCGCGCATG 4132224
QY 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 4132223 GAGTGGCGCGCGCAGCAGCGCGCGCGCTGCTCACTGAGCTGCGCGCGCGCGCGCACACC 4132164
QY 123 gileHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsnSerArgG1 143
Db 4132163 CCGAGATCGACCGCGCTGGAA-GCGAGGTCAACAAGCTC----- 4132125
QY 143 nValAspGluTyValArgAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 4132124 -----TCCAGGAGAGGCGATCTCTTCGCGATCGCGCGCGCAA 4132084
QY 161 nGluGlyPro---AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
Db 4132083 CGAAGCGGAGTTCGGCGAGCAGACCATCGCTCCCGCGCAGCGCGCGCGCGCTCAC 4132024
QY 180 rValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHi 200
Db 4132023 CGTCGGCGCC-----GTGAACGACGCGCAA 4131997
QY 200 sileAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220
Db 4131996 GCTGGCGTCTCTTCCAGCGCGCGCGCGCTTGGAGCGCGCATCAAGCCGCGCTCAC 4131937
QY 220 rAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTr 240
Db 4131936 CGCACCGCGCGTGACATCACCGCGCC-----GCGCGCGCGCGCGCGCTCATCGA 4131886
QY 240 pAlaAsnTyArgSerLys-----TyAlaTyMetGlyGlyThrSerMetAlaTh 257
Db 4131885 CCAGGAGTCCGCGCAGAAAGCGCGACCGGTACCTCACCATCTCCGTACGTCGCGCG 4131826
```

```
QY 257 rProfileValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyI 277
Db 4131825 CCGCATGTCGGGGCGCGCGGATCTCAAGCAGCAGCAC----- 4131783
QY 277 eThrProLysProSerLeu-----IleLysAlaAlaLeuIleAlaGlyAlaThrAspVa 295
Db 4131782 ----CCCAACTGCTGTCGCGAGCTCAAGGCGCGCTGACCGGCTCGCGAAG----- 4131732
QY 295 lGlyLeuGlyTyProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLe 315
Db 4131731 -GGCGCAAGTACACCGCTTCAGCAGGCGCTCGGCGCTATCGCGCGCAGCGCAT 4131673
QY 315 uAsnValAlaTyValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTySe 335
Db 4131672 CRAAGCATGCTGATCGCAACCGAAGCTCGGTGAGCTTCGGCAT-CCAGCAGTGGCGGC 4131614
QY 335 :PheGlnThrGlnAlaGlyLysPro 343
Db 4131613 ACACCGACGACAGCGCGGTCAACCA 4131589

RESULT 11
US-10-090-624-5
; Sequence 5, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOUO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 1519693/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 4765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-5

Alignment Scores:
Pred. No.: 1,37e-22 Length: 4765
Score: 305.00 Matches: 128
Percent Similarity: 36.26% Conservative: 37
Best Local Similarity: 28.13% Mismatches: 154
Query Match: 13.48% Indels: 136
DB: 13 Gaps: 15

US-09-985-689a-3 (1-433) x US-10-090-624-5 (1-4765)
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1024 GGCAATGTTATGACATTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1083
QY 41 -----SerMetHisGluAlaPheArgGly 48
Db 1084 GTTCCACTTGGCCAGTCAAGTTACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1143
QY 49 LysIleThrAlaLeuTyThrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn----- 66
Db 1144 CCTCTCAAC-----TACGTGCTTCAGAAATA-----GATCTCAACGGAGAA 1185
QY 67 -----GlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1186 TATGCAGTATTGGGTGGGTGGTGCACGGTCAACGGTCAACGGTCAACGGTCAACGGTCAACGGT 1245
```

```
QY 78 GlyAsnAlaLeuAsn----- 82
Db 1246 GGTTACGACGACCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1305
QY 83 -----LysGly 84
Db 1306 GTGTTCCTCAAGACTCTATGTTGGGATATATCAAGAGTTTACCACAGACACCCTGGTCAGGGT 1365
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 104
Db 1366 GTTGCTCCAGGTGCCCAATAATAGGCAATAAGAGTTCTTAGGAGTGTATGGA-----CGG 1419
QY 105 GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIle 124
Db 1420 GGTAAGCATGTTGGATATTATAGAAAGTATACATACGCAACCAACCCATGTTGACAGAGTT 1479
QY 125 HisThrAsnSerTrpGly-----AlaProValAsnGlyAlaTyThrAlaAsnSerArg 142
Db 1480 ATAAGCATGAGTCTCGGTGGAAATGCTCCATCTTAGTGGTACTGATCCAGAAAGCGTT 1539
QY 143 GlnValAspGluTyValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 1540 GCTGTGATGAGTTACCGGAAAGTACGGTGTGTATTCGTAATAGTGCAGGAATGAA 1599
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 1600 GGTCTGGCATTAACATCGTTGGAGTCTCTGGTGTTCGCAACAAAGGCAATACTGTTGGA 1659
QY 183 AlaThrGlu----- 185
Db 1660 GCTGTGCTGAGTCCCATTAACGTTGGAGTTATGTTTCCCAAGCACTTGGATATCTCTGAT 1719
QY 186 -----AsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHisIle 201
Db 1720 TACTATGATCTTATTACTTCCCGCTACACAACGTT-----AGAATA 1764
QY 202 AlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAla 221
Db 1765 GCATTCTCTCAAGCAGAGGCGCGAGAATAGATGTTGAAATAAAACCAATAGTAGTGCT 1824
QY 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 241
Db 1825 CCAGTTACGGAAATTACTATCCCTCCGATGGGATGGGAGTGGCGGAGCTGACTTC----- 1878
QY 242 AsnTyArgSerLysTyArgAlaTyMetGlyGlyThrSerMetAlaThrProIleValAla 261
Db 1879 -----ATGCTGGAACCTTCGATGGCTACTCCACATGTGCTCAGC 1914
QY 282 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 281
Db 1915 GGTGTCGTGTCACCTCCCTCATAGCGGG---GCAAGGCGCGGGAATATATACTACAATCCA 1971
QY 282 SerLeuIleLysAlaAlaLeuIleAlaGlyAlaThr-----AspVal 295
Db 1972 GATATAATTAAAGAGTTCTTGAGAGGCGTGCAACCTGGCTTGAGGAGATCCATATACT 2031
QY 296 GlyLeuGlyTyProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeu 315
Db 2032 GGGCAGAAAGTACACTGAGCTTGACCAAGGTCATGGTCTTGTAACTGTTACCAAGTCTCTGG 2091
QY 316 AsnValAlaTyArgValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTySer 335
Db 2092 GAAATC----- 2097
QY 336 PheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGly 355
Db 2098 CTTAAGGCTATAAAGCGGACCACTCTCCCAATTGTTGATCACTGGGAGCAAGTCTCTAC 2157
QY 356 SerThrThrAlaSerTyThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsn 375
Db 2158 AGCGACTTGGGAGTACTTGGGTGTG---GACGTTATAAGAGGTCTCTACGCAAGGAC 2214
```



```

; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1.
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:
Pred. No.: 3,398-22 Length: 1560
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservative: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.04% Indels: 90
DB: 15 Gaps: 19

US-09-385-689A-3 (1-433) x US-10-084-846A-113 (1-1560)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAlaSerThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 451 GGGCAGGGAGTGACGGCGTACGTCTATCGACACCGCGTC-----CGC 492
Qy 41 SerMetHisGluAlaPheArgGlyValThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 493 ATCACCCACAGCGACTTCGGCGCGCGGCGCTCTACGGCTACGACGCCATCGACAACGAC 552
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 553 AACACCGCCAGCAGCGCGCCACCGCCACGCGACGCGACGTCGCCGGCAGCGTCCCGGCAAC 612
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 613 GCCTAC-----GGCGTCGCCAAGAAGGCCAAGATCGTAGGGCTCGCGTGTGAACAAC 666
Qy 100 Ser-----GlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 667 TCCGCCAGGGGACCACCGCCAGGTGTCGCCGCGCATCGATCGGTCTGCCGGAACGCC 726
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 727 GT-CAAGCGCGCGTCGCCAAT-----GTCCCTCGCGGGCGGCCCGA 770
Qy 137 rThrAlaAsnSerArgGlnValAspGluTyrValArgAsn-----AsnAspMetThr 154
Db 771 CACGGCC-----CTGCACACGGCCGTACGCAACGGCATCGCTCGCGCGTCAC 818
Qy 154 rValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db 819 CTTCCCGCTGGCGCGCGCAACAGTCGACCAACGGCTCCACGAGTCA---CCCGCAGC 875
Qy 174 rAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyArgProSerPheGlySerIl 194
Db 876 CGTCCACCGAGGCCATCACGGTCGGCGCAGCACACGACGTCGGACGCCCAAGCCCGCC 930
Qy 194 eAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 931 -----TACTCCAACTACGGTCCGTCCTC----- 954
Qy 214 gIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db 955 -----GACCTCTCGCCCGCGGTTCGTCCATCACCCTCGGCC----- 990
Qy 234 aProAspSerSerPheTrpAlaAsnTyArgAsnSerLysTyArgAlaTyMetGlyGlyThrSe 254
Db 991 -----TGGNACTCAGCGACTCGCGCACCAACACCATCTCCGTGACGTC 1034

```

Db 56648 GGGCAGGAGTACGGCGTACGTATCATCGACACCGCGCTC-----CGC 56689
Qy 41 SerMetHisGluAlaPheAArgGlyLeuThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 56690 ATCACCACACGCGACTTCGGCGCGCGCGCTCTACCGGTACACGCCATCGAACACGAC 56749
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 56750 AACACCGCCAGAGCGCCACGGCCACGCGACGACGCGTGGCCGCGACGCTCGCGCGAAC 56809
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 56810 GCCTAC-----GGCGTCGCCAAGAACGCCAAGATCGTAGGCGTCGCGGTGTAACAAC 56863
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrIlePheSerGlnAla 117
Db 56864 TCCGGCAGGCGACACCGCGCGAGTCTGTCGGCGATCGACGTGGTGGCGCGGACGCGC 56923
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 56924 GT-CAAGCGCGCGTCCCAACAT-----GTCCCTCGCGCGCGCGCGCA 56967
Qy 137 rThrAlaAsnSerArgGlnValAspGluTyrValArgAsn-----AsnAspMetTh 154
Db 56968 CACGGCC-----CTCGACACGGCGGTACGCAACGCCATGGCGCTCCGGCGGTAC 57015
Qy 154 rValLeuPheAlaAlaGlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db 57016 CTTCCGCGTGGCGCGCGACGAGTGCACCAACGCGCTCCACGAGTCA---CCCGCAGC 57072
Qy 174 rAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIl 194
Db 57073 CGTCACCGAGGCCATCAGTGGCGCGCGACGACGCTCGGACGCCAAGCGCGCGC----- 57127
Qy 194 eAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 57128 -----TACTCAACTACGGCTCCGCTC----- 57151
Qy 214 gIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerLeuAl 234
Db 57152 -----GACCTCTTCGCGCGCGGTCTGCTCATCATCCTCGGCG----- 57187
Qy 234 aProAspSerSerPheTTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSe 254
Db 57188 -----TGAAGTCAAGGAGTCTCGGACCAACACCATCTCCGTACGTC 57231
Qy 254 rMetAlaThrProfileValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
Db 57232 GATGGCGACCGCGACGTCGGCGCGCGCGCGCTC-----CACCTCGCGCGCAA 57282
Qy 274 nArgGlyIleThrProLys-----ProSerLeuIleLysAlaAlaLeuIleAlaGl 291
Db 57283 CCCCTCGGCGCCCGCTCCAGGTCCGACCGCGCTGACGTCCGCGCGCCACACCGCGCT 57342
Qy 291 yAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLe 311
Db 57343 CGTCACCAACCCCGCGACGCGCTCGCCCAAC----- 57373
Qy 311 uAspLysSerLeuAsnValAlaTyValAsnGluAlaThrAlaLeuThrThrGlyGlnLy 331
Db 57374 -----CGGCTCTGTAGTCTGCGCGCGCGCGACACCGCTCCGCGCGCGCG 57420
Qy 331 s-----AlaThrTyrSerPheGlnThrGlnAlaGlyLysProLeuLy 345
Db 57421 CTTTCGAGAACCCGCTGACTACCATCATCGGACAACTCCACGCTCGAGTCCCGGTGAC 57480
Qy 345 sIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
Db 57481 GGTCTTCGCGGTCTCCGCGCAACGCGCGCTCGCGCGCTCCGCGGTAGAGGTCCACATCGTCCA 57540
Qy 361 rThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGl 381
Db 57541 CACGTACATCGCGGACCTTCAGGTCCAGTGTATCGCCCCCGACGCGCGGTACAGCT 57600

Qy 381 yAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVa 401
Db 57601 CAAG-----TCGTACGGCACCGCGCGAGTTCGGAACAATCAACACCATCTCGT 57654
Qy 401 lPheIleAsnAlaProGln-----SerGlyThrTyrThrIleGluVal 415
Db 57655 G-----AACGCTCTCGGAGCGCGCCACGCGACGTCGGAACTCGCGGTG 57700

RESULT 15

US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNENEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

Alignment Scores:
Pred. No.: 6,68e-20 Length: 59816
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservative: 51
Best Local Similarity: 25.90% Mismatches: 153
Query Match: 13.04% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-3 (1-433) x US-10-084-846A-2 (1-59816)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGAGTACCGCGTACGTATCGACACCGCGCTC-----CGC 3128
Qy 41 SerMetHisGluAlaPheAArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 3127 ATCACCACACGCGACTTCGGCGCGCGCGCTCTCTACGCTACGCGCATCGACACGAC 3068
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 3067 AACACCGCCAGAGCGCGCCACCGCGCACGCGACGCGTGGCGCGCGACGCTCCCGCGAAC 3008
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 3007 GCCTAC-----GGCGTCGCCAAGAACGCCAAGATCGTAGGCGTCCGCGTGTGAACAAC 2954
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrIlePheSerGlnAla 117
Db 2953 TCCGGCAGGCGCACACCGCGCGAGTCTGTCGGCGCATCGACTGGTCCCGCGGACGCGC 2894
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 2893 GT-CAAGCGCGCGTCCCAACAT-----GTCCCTCGCGCGCGCGCGCA 2850
Qy 137 rThrAlaAsnSerArgGlnValAspGluTyrValArgAsn-----AsnAspMetTh 154
Db 2849 CACGGCC-----CTCGACACGGCGCTACGCAACGCCATGGCGCTCCGGCGGTAC 2802
Qy 154 rValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyTh 174

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2274.58 Seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-3
Perfect score: 2263
Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVPSGPQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-Q=/cgn2.1/USPTO.spool/US09985689/runat.31032004.161807.4168/app_query.fasta_1.3498
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689@cgn.1.1.7084/runat.31032004.161807.4168 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	235.5	10.4	1605	13	BO622771	BO622771 CC Contig
2	228	10.1	4198	11	AK029048	Mus muscu
3	214.5	9.5	640	12	BJ395336	BJ395336
4	213.5	9.4	594	12	BJ393752	BJ393752
c	204	9.0	2141	13	BQ143519	BQ143519
6	200.5	8.9	532	29	TA319G10P	TA319G10P
7	199.5	8.8	718	12	BI750157	BI750157
8	195.5	8.6	771	14	CA320325	CA320325
9	193.5	8.6	601	12	BJ387574	BJ387574
10	190	8.4	545	13	BQ143519	BQ143519
11	188	8.3	574	29	TA319G10P	TA319G10P
12	188	8.3	614	9	AJ273402	AJ273402
c	185.5	8.2	2121	28	B2424995	B2424995
14	182.5	8.1	665	13	BQ770462	BQ770462
c	182.5	8.1	716	28	B2893395	B2893395
c	180	8.0	530	29	CNS010PO	CNS010PO
17	180	8.0	650	9	AJ274038	AJ274038
c	179.5	7.9	675	14	CF72824	CF72824
19	179	7.9	508	28	AQ652212	AQ652212
c	178	7.9	633	12	BJ369190	BJ369190
21	177.5	7.8	576	14	CD295943	CD295943
22	177.5	7.8	1002	29	CNS06D6B	CNS06D6B
23	176.5	7.8	594	14	CF846026	CF846026
c	176.5	7.8	1029	29	CNS071DW	CNS071DW
25	175	7.7	580	9	AJ273745	AJ273745
26	175	7.7	583	9	AJ273947	AJ273947
27	175	7.7	593	9	AJ273918	AJ273918
28	175	7.7	601	9	AJ273050	AJ273050
29	175	7.7	601	9	AJ273921	AJ273921
30	173.5	7.7	641	12	BJ393925	BJ393925
31	172	7.6	610	9	AJ274218	AJ274218
c	170	7.5	584	28	AQ651427	AQ651427
33	169.5	7.5	781	13	BQ612128	BQ612128
34	169.5	7.5	937	12	BG246418	BG246418
35	169	7.5	609	9	AJ272712	AJ272712
36	169	7.5	794	14	CF737198	CF737198
37	167.5	7.4	604	9	AJ274126	AJ274126
38	167.5	7.4	615	9	AJ274059	AJ274059
39	167	7.4	616	14	CF138007	CF138007
40	167	7.4	681	14	CB690041	CB690041
41	166.5	7.4	895	13	BQ216158	BQ216158
42	166	7.3	619	9	AJ273097	AJ273097
43	166	7.3	887	13	BQ879057	BQ879057
44	165.5	7.3	593	9	AJ273903	AJ273903
45	165.5	7.3	596	9	AJ273379	AJ273379

ALIGNMENTS

RESULT 1
BO622771
LOCUS
DEFINITION
CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
cdna, mRNA sequence.
ACCESSION
BO622771
VERSION
BO622771.1
KEYWORDS
EST.
SOURCE
Conidiobolus coronatus (Delacroixia coronata)
ORGANISM
Conidiobolus coronatus
Eukaryota; fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE
1 (bases 1 to 1605)

BO622771
CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
cdna, mRNA sequence.
ACCESSION
BO622771
VERSION
BO622771.1
KEYWORDS
EST.
SOURCE
Conidiobolus coronatus (Delacroixia coronata)
ORGANISM
Conidiobolus coronatus
Eukaryota; fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE
1 (bases 1 to 1605)

BO622771
CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
cdna, mRNA sequence.
ACCESSION
BO622771
VERSION
BO622771.1
KEYWORDS
EST.
SOURCE
Conidiobolus coronatus (Delacroixia coronata)
ORGANISM
Conidiobolus coronatus
Eukaryota; fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE
1 (bases 1 to 1605)


```

QY 314 SerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThr 333
   |||
Db 1827 GCTTATCAGATC-----CTC 1841
   |||
QY 334 TyrSerPheGlnThrGlnAlaGly---LysProLeuLysIleSerLeu----- 348
   |||
Db 1842 AGCAGCTATAAACCCGACGAGCCTGAGTCTAGCTACATCGACCTGACTGAGTGTCCC 1901
   |||
QY 349 ---ValTyrThrAspAlaProGlySerThrThrAlaSerTyr-----Thr 362
   |||
Db 1902 TACATGTGG-----CCCTACTGCTCCCGACCTACTCTACTATGGAGGAATGCCAAC 1952
   |||
QY 363 LeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsn 382
   |||
Db 1953 ATCGTTAT-----CTCACCATCCTCAATGGCATGGCGGTCCACAGGAAGA 1997
   |||
QY 383 AspPheSerTyrProTyrAspAsnAsnTrp-----AspGlyValAsnAsn 397
   |||
Db 1998 ATTGTGGATAGCCT-----GAGTGGCGACCCCTATTTACCACAGAAATGGAGACAA 2048
   |||
QY 398 ValGluAsnValPhe-----IleAsnAlaProGlnSerGlyThrTyrThrIle 413
   |||
Db 2049 ATTGAAGTGGCTTCTCCTACTCCTCAGTGTGTGGCCCTGGTCAGGTACCTTGGCCATC 2108
   |||
QY 414 GluVal 415
   |||
Db 2109 TCATT 2114

RESULT 3
LOCUS BJ395336
DEFINITION BJ395336 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds38b16 5', mRNA sequence.
ACCESSION BJ395336.1 GI:19306422
VERSION BJ395336
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 640)
AUTHORS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
Tadaasu Shin-i, Tanaka, Y., Kohara, Y. and Shin-i, T.
JOURNAL Full length cDNA of Dictyostelium discoideum at the slug stage
COMMENT Unpublished (2002)
Contact: Tadaasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..640
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds38b16"
/sex="mat A"
/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

FEATURES
source
Pred. No.: 2,44e-11 Length: 640
Score: 214.50 Matches: 67
Percent Similarity: 50.47% Conservative: 41
Best Local Similarity: 31.31% Mismatches: 65
Query Match: 9.48% Indels: 41
DB: 12 Gaps: 12

US-09-985-689A-3 (1-433) x BJ395336 (1-640)

```

```

QY 19 LeuTyrGlyGlnGlyGlnValValAlaAspThrThrGlyLeuAspThrGlyArg--- 37
   |||
Db 22 TTAAGAGGTAAAGGTACAGATATTGAGTATTGCTGATCTGTTAGTGGTAGCCATTGT 81
   |||
QY 38 ---AsnAspSer-----SerMetHisGluAlaPheArgGly 48
   |||
Db 82 TTCTTTTCAGATTCAAAGTATCCATATCAATTTATCAAGTGAATGAATAATCATAGAAA 141
   |||
QY 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspPro-----Asn 66
   |||
Db 142 GTTGTAACTTATATT-----ACTTACCATGACAATGAAGATTATGTAAAT 186
   |||
QY 67 GlyHisGlyThrHisValalaglySerValLeuGlyAsn-----Ala 80
   |||
Db 187 GGTCAATGGTACACATTTTGGCTCTGCAGCAGGTACTCCAGAGANTTCTTCATGGGCT 246
   |||
QY 81 LeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98
   |||
Db 247 ATTTCATCATTTAGTGTCTTCACTGACGCAAGAGATTGCTTTTATGATCTT----- 300
   |||
QY 99 SerSerGlyGlyLeuGlyLeu---ProSerAsnLeuAsnThrLeuPheSerGlnAla 117
   |||
Db 301 TCATCTGGAAGTTCTGAACCAACACACCCGAGATTACAGTCAATGTACAAACCATTA 360
   |||
QY 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----ProVal 133
   |||
Db 361 TATGATGCAGTGCACAGAGTACATGCTGATTTCTTGGGGTCTGTATCTTTGCCAAGTTAT 420
   |||
QY 134 AsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrVal---ArgAsnAsnAsp 152
   |||
Db 421 TATGCTGGTTATTCCTGATGATGCTGGTGTATGATGCTTCTCTATGAGTACCCAGAA 480
   |||
QY 153 MetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaPro 172
   |||
Db 481 TTCTTATATTAAGAGCTGCTGTAT---AACGAGCTATTTGCACTTTATAGCTCAA 537
   |||
QY 173 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGly 192
   |||
Db 538 GCAACAGCTAAATGCAATTACAGTTGGTGTGCTGAGCAACACAGCTCATGTAAATTATGTG 597
   |||
QY 193 Ser-----IleAlaAspAsnProAsn 199
   |||
Db 598 TCAGATGCATTGGATATTATGATTCTTCAGATAATGCTAAT 639
   |||

RESULT 4
LOCUS BJ393752
DEFINITION BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds32b16 5', mRNA sequence.
ACCESSION BJ393752.1 GI:19304838
VERSION BJ393752
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 594)
AUTHORS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
Tadaasu Shin-i, Tanaka, Y., Kohara, Y. and Shin-i, T.
JOURNAL Full length cDNA of Dictyostelium discoideum at the slug stage
COMMENT Unpublished (2002)
Contact: Tadaasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..594
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds32b16"

FEATURES
source

```

```

/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="nictovostelium discoideum cdna library. SP"

```

ORIGIN

Alignment Scores:		
Pred. No.:	2.5e-11	Length: 594
Score:	213.50	Matches: 54
Percent Similarity:	31.48%	Conservative: 38
Best Local Similarity:	31.21%	Mismatches: 60
Query Match:	9.43%	Indels: 21
DB:	12	Gaps: 8
		MS-09-985-689A-3 (1-433) X BU393752 (1-594)

```

Qy 255 MetAlaThrProIleValAlaGlyAenValAlaGlnLeuArgGluHisPheIleLeuAsn 274
Db 471 ATGGCTACTCCCATATTCTGGT-----CTTGTGCTACTTTCAGTGCTCTC 424
Qy 275 ArgGlyIleThrProIleValAlaGlyAenValAlaGlnLeuArgGluHisPheIleLeuAsn 294
Db 423 ACGGCAAGACTAGCCCTCCGCTCTTTCCAGAGATCCAGGACACTTCTACCAAGAAC 364
Qy 295 ValGlyLeuGlyTyrProSerGly 302
Db 363 GTGATCCGCAATGTGCCGCTGTC 340

RESULT 6
TA319G10P 532 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 319g10, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL492464
VERSION AL492464.1 GI:11867408
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
REFERENCE 1 (bases 1 to 532)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhi@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTAT 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.
FEATURES
source
1..532
Location/Qualifiers
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="319g10"

ORIGIN
Alignment Scores:
Pred. No.: 4,45e-10 Length: 532
Score: 200.50 Matches: 58
Percent Similarity: 46.20% Conservative: 27
Best Local Similarity: 31.52% Mismatches: 72
Query Match: 8.86% Indels: 27
DB: 29 Gaps: 5

US-09-985-689A-3 (1-433) x TA319G10P (1-532)

Qy 92 ValPheGlnSerIleMetAspSerSerGlyGlyLeuGly----- 104
Db 2 GTAATGCCAAGTATAATGCTGCCAGGGGGGGAAGATTCTTCAGGGGTTGGGCTGCC 61
Qy 105 -----GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTTP 118
Db 62 CATCCAGTCAGACTTGCTCTCCGCCAGCAGCTACTCAAAATTATTCGCCGGTATAT 121

```

```

Qy 119 AsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 138
Db 122 GGCGCTGGAGCCCGTGTCTCTCAAACTCGTGGGGTTTGTCTCCCTCCGAGTATCT 181
Qy 139 AlaAsnSerArgGlnValAspGluTyrValArgAsn---AsnAspMetThrValLeuPhe 157
Db 182 CTTGTGGAAAAGATATGGATGGTTTGGAGTAGTATGACGATGGCTACTTATCTTC 241
Qy 158 AlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLeuAsn 177
Db 242 TCCACTGGCAACAGTATCCAAAGATGCG-----CTAATGACTCCGCTCGTGGTAAAGAAC 295
Qy 178 AlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsn 197
Db 296 GTGATGTGCGTGGGTCACACAAAAACGTTGTGACGCTTCGAAAACACTT----- 346
Qy 198 ProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLeuPro 217
Db 347 -----GTTTCTTCGTTTCTTCGCGATGTCCTCAACATACGACGGTAGGATGAACCC 397
Qy 218 AspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 237
Db 398 GATCTTGTGCGTCCGCGGAGAGGTCGTCTGCTCTTCTCTCTGGCAAAGCATCAGCT 457
Qy 238 SerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 257
Db 458 AAA-----CAATGTAAGTGGTGGCCAAAGCGGNGTTTCATCGATGGCAACT 502
Qy 258 ProIleValAla 261
Db 503 GCGGCGCTGCGG 514

RESULT 7
BI750157 718 bp mRNA linear EST 25-SEP-2001
LOCUS Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence.
DEFINITION Fg02_10g08, Fg02_AAFC_ECOEC_Fusarium graminearum mycelium
Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence.
ACCESSION BI750157
VERSION BI750157.1 GI:15771959
KEYWORDS EST.
SOURCE Gibberella zeae
ORGANISM Gibberella zeae
REFERENCE 1 (bases 1 to 718)
AUTHORS Harris, L.J., Glassco, T., Rocheleau, H., Allard, S., Chapados, J.,
Couroux, P., De Moors, A., Hattori, J.I., Ouellet, T., Robert, D.S.,
Singh, J.A., Spott, D. and Tinker, N.A.
Expressed Sequence Tags from Fusarium graminearum mycelium
Unpublished (2001)
Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris@em.agr.ca
Location/Qualifiers
1..718
/organism="Gibberella zeae"
/mol_type="mRNA"
/strain="DAOM 180378"
/db_xref="taxon:5518"
/clone="Fg02_10g08"
/tissue_type="mycelial tissue"
/dev_stage="Asexual"
/clone_lib="Fg02_AAFC_ECOEC_Fusarium graminearum mycelium"
/site="Bluescript SK+/XhoI-ECORI; Site 1: EcoRI;
Site 2: XhoI; Mycelial tissue was collected from v8 agar
plates after a growth period of 6-7 days at 25 C with 14
hrs (FL/UV) day lighth exposure. Mycelia was ground in

```

liquid nitrogen prior it's storage at -80 C until RNA extraction. Directional cloning with 5' end of cDNA cloned into EcoRI site of pBluescript and 3' end of cDNA cloned into XhoI site of pBluescript (Stratagene, La Jolla, CA)."

ORIGIN

Alignment Scores:
 Pred. No.: 9.44e-10 Length: 718
 Score: 199.50 Matches: 70
 Percent Similarity: 44.20% Conservative: 29
 Best Local Similarity: 31.25% Mismatches: 84
 Query Match: 8.82% Indels: 41
 DB: 12 Gaps: 9

US-09-985-689A-3 (1-433) x B1750157 (1-718)

Qy 46 PheArgGlyLysIleThrAlaLeuTyxAlaLeuGlyArgThrAsnAlaAsnAspPro 65
 Db 9 TTCGAGGTCGTCTCAGGCTGTCTACACTCCCTCAGCGCCAGACGCT--GACACC 65

Qy 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsnLysGlyMet 85
 Db 66 AACGGTCACGGAACCTACGTTGCTGGCACTATTCCCGAAGACATAC-----GGTGT 119

Qy 86 AlaProGlnAlaAsnLeu-----ValPheGln---SerIleMetAspSerSer 100
 Db 120 GCCAAGAGGCCACCATCAGCTGTCCAGGTCCTCCAGGTCAGTTCATCCAGCACCTCC 179

Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
 Db 180 ATCATCTCTCGTGGCTTCACTGGGCTGCCAACACACATCATCTCCAAAG----- 227

Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsn 140
 Db 228 ---GCCGAACACAGACTCAGTCGTCATATGCTCTCGCGGTGGTACTCTGCTCTCC 284

Qy 141 SerArgGlnValAspGlyLysValArgAsnAspMetThrValLeuPheAlaAlaGly 160
 Db 285 TTCACCAACGCTGTCAGTCTKCTCCAGCTCCGCTATTATCTCTGCCATTGTCGCGGT 344

Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 345 AACGATGTGGCAACGCTGCCAACACTTCT---CTGCCTCTCTCCCGCCGCAWCACT 401

Qy 181 ValGlyAlaThrGluAsnTyxArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
 Db 402 GTCGTGCC-----ATTGACAGCACTGGGCC 428

Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 429 ATTCTCTGTACTCCAACTACGTCACCGTCTCTC-----GATCTCTTT 470

Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 Db 471 GCYCTKGACACGAGTCTCTCCGCC-----TGG 500

Qy 241 AlaAsnTyxAsnSerLysTyxAlaTyxMetGlyThrSerMetAlaThrProIleVal 260
 Db 501 TACACGACGACAGTKCCACCAACACCATCAGCGGKACSTCCATGGCTACTCYCCACATT 560

Qy 261 AlaGlyAsnVal 264
 Db 561 GCGGACTGTCTC 572

RESULT 8
 CA320325
 LOCUS
 DEFINITION
 CA320325
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE

CA320325
 UI-N-FWO-cby-g-23-0-UI-r1 NIH_BMAP_FWO Mus musculus cDNA clone
 IMAGE:6816072 5', mRNA sequence.
 CA320325.1 GI:24538449
 EST.
 Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 771)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgsbbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pyX-5.

FEATURES

Location/Qualifiers

source

1..771
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CS7BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6816072"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_FWO"
 /note="Organ: Brain; Vector: pyX-Asc; Site:1: EcoR I;
 Site:2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pyX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCAGACAG. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 Program coordinator."

ORIGIN

Alignment Scores:
 Pred. No.: 2.74e-09 Length: 771
 Score: 195.50 Matches: 78
 Percent Similarity: 40.21% Conservative: 37
 Best Local Similarity: 27.27% Mismatches: 114
 Query Match: 8.64% Indels: 57
 DB: 14 Gaps: 10

US-09-985-689A-3 (1-433) x CA320325 (1-771)

Qy 8 ValIysAlaAspValAlaGlnAsnAsnTyxGlyLeuTyxGlyGlnGlyValValAla 27
 Db 40 CTCGAGGACATGTGTGTGG---CAGATGGGATACACAGGTGCTAATGTCAGATTGCT 96
 Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 97 GTTTTGTACTGGCTC-----AGTGAGAAGCATCCGCATTTTAAG 138

Qy 48 GlyLysIleThrAlaLeuTyxAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
 Db 139 AAT-----GTGAAGGAGAGAACCACTGGACCAATGAGCGGACCTCG 180

Qy 64 ---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn 82
 Db 181 GATGATGGGTAGGCCATGCGACATCGTTGCGAGGTGTGATTGCCACGATGAGGGAGTGC 240

```

QY 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly 101
Db 241 CAAGAGATTGCTCCAGAGATGAGAGCTGCACATCTTCAGGGCTCTTACCAACAATCAG--- 297
QY 102 GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGly 121
Db 298 : : : : : GTGCTTTACACATCTGGTTCTCGGATGCGCTTCAACTATGCC 339
QY 122 AlaA-GilleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsnSer 141
Db 340 ATCTTAAGAAGATGACGCTCTCAACCTTAGCATCGGTGGCCGCGACTTCATGGATCAT 399
QY 142 ArgGlnValAspGluTyR-----ValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 400 CGGTTTGTGACAAAGGTGGGAATTAACAGCTAACTGAATGAATATGGTTCTCGCTATT 459
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIleAsnAlaIle 179
Db 460 GGCAATGATGGACCTCTCTATGGCACTCTGAATAACCCCTGCTGATCAGATGGATGATT 519
QY 180 ThrValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 520 GGAGTGGGTGGC-----ATTGACTTTGAAGAT 546
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArg-----AspGly 213
Db 547 ACATCGCTCGCTTTCTCCAGGGGAATGACTACTGGGATTAACAGAGGCTATGGT 606
QY 214 ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerLeu 233
Db 607 CGTGTGAAGCCTGACATTTGCTACCTATGCTGTGGTGTGGAGTGGGGGT----- 651
QY 234 AlaProAspSerSerPheTrpAlaAsnTyArgSerLysTyThrAlaTyMetGlyGlyThr 253
Db 652 : : : : : TCGGTGTGAAGGGGGTGGCTGGCTGCTCCTCAGGGACC 690
QY 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheLeuLys 273
Db 691 AGTGTGCTTCCCAAGTGGTGTGGTGGGGCGTGCACCTTGTAGTANGCACAGTACAGAAG 750
QY 274 AsnArgGlyIleThrPro 279
Db 751 CGGAGCTGGTGAATCCT 768

RESULT 9
BJ387574 601 bp mRNA linear EST 08-MAR-2002
LOCUS Dictyostelium discoideum cDNA library, SF Dictyostelium
DEFINITION dictyostelium cDNA clone dds3a18 5', mRNA sequence.
ACCESSION BJ387574
VERSION BJ387574.1 GI:19296958
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE 1 (bases 1 to 601)
AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. .601
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clones="dds3a18"

FEATURES
source

```

```

/sex="mat A"
/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

```

ORIGIN

```

Alignment Scores: 2.87e-09 Length: 601
Pred. No.: 139.50 Matches: 55
Score: 139.50 Mismatches: 21
Percent Similarity: 46.34% Conservative: 49
Best Local Similarity: 33.54% Indels: 39
Query Match: 8.55% Gaps: 6
DB: 12

US-09-985-689a-3 (1-433) x BJ387574 (1-601)

QY 197 AsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLys 216
Db 51 AATGGAATAAATATTGTCATCTCATCAAAGGTCACACATCATGCTAGATGAAA 110
QY 217 ProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla----- 234
Db 111 CTGATTTAGTTCGCCCTCGTGAATATATTACATCGCAAGATCAATAGTGCCAAATACA 170
QY 235 -----ProAspSerSerPheTrpAlaAsnTyArgSerLysTyThrAlaTyMet 250
Db 171 ACAGACCAATGTGGTGTGCTCTTTA-----CCAAATACAAATGCATTTATGGCG---ATA 224
QY 251 GlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHis 270
Db 225 TCTGGTACATCAATGGCACTCTCATTTGCAGCAGCAGCAACAATCTCTTAGACAATAT 284
QY 271 -----PheIleLysAsnArgGlyIleThrProLys 280
Db 285 TTAGTTGATGTTATTATCAACTGTTTCAATTGTAGATCAATAAATTAACAACCAACT 344
QY 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAla----- 292
Db 345 GGATCATTTTAAAGCAATTAATGATTAATTAATGCTCAGTATTATTAATGCTACATTTCAA 404
QY 293 -----ThrAspValGlyLeuGlyTyProSerGlyAsp----- 303
Db 405 TTGATTATCATCATCAAGTATTATCATATCCATCAACAACCAAGTTTTGAATAATTTGCAGGT 464
QY 304 -----GlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyR 319
Db 465 GCAAGTTTAGTTCAAGTTGGGGTGGTATTAGATGAGTAATGGTTACATGTTGTCAT 524
QY 320 ValAsnGluAla 323
Db 525 AATAATAATAGT 536

RESULT 10
BJ387574 545 bp mRNA linear EST 17-SEP-2002
LOCUS TgSTzyb80b07.y1 TgVEG Partially sporulated oocyst cDNA Toxoplasma
DEFINITION gondii cDNA clone TgSTzyb80b07.y1 5', similar to TR:Q45522 Q45522
ACCESSION BUS75479
VERSION BUS75479.1 GI:23045230
KEYWORDS EST.
SOURCE Toxoplasma gondii
ORGANISM Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
REFERENCE 1 (bases 1 to 545)
AUTHORS Tang, X., Cole, R., Fogarty, S., Sibley, L.D., Ajioke, J.A., White, M.,
Clifton, S., Fape, D., Martin, J., Wylie, T., Dante, M., Marra, M.,
Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M.,
Ritter, E., Bennett, J., Franklin, C., Tsagareishvili, R., Ronko, I.,
Kennedy, S., Maguire, L., Waterston, R. and Wilson, R.
TITLE Toxoplasma EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Clifton, S.

```

Toxoplasma EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@wustl.edu
Contact David Sibley (toxoeast@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 445.

Location/Qualifiers
1. 545
/organism="Toxoplasma gondii"
/mol_type="mRNA"
/strain="VEG (Type III)"
/db_xref="taxon:5811"
/clones="TG85ryb80b07.yl"
/dev_stage="Partially sporulated oocysts"
/clone_lib="TGVEG Partially sporulated oocyst cDNA"
/note="Vector: Modified pBluescript (pBSSK+); Site 1:
BamHI; Site 2: EcoRI; PolyA mRNA from partially-sporulated
oocysts was converted to cDNA using the
template-switching PCR method (SMART cDNA, Clontech Inc.,)
and sized selected on SizeSep 400 columns (Amersham
Pharmacia Biotech Inc.). First strand was reverse
transcribed using the CDS III-oligo-dT primer and a 5'
template switch primer (Smart IV primer). The product of
the first strand synthesis was PCR amplified using the
same primer set and the fragments were digested with SfiI.
The fragments were size selected, ligated into a modified
pBluescript vector containing directional SfiI sites, and
electroporated into DH10B or DH12S cells. Vector: SfiI
sites were added to the multiple cloning region of
pBluescript SK+ between the BamHI/EcoRI sites. The
modified polylinker has the following sequence:
5'-GATTCGGCATTCAGCGCC(G)n--insert--
GCCGCCCTCGCCACGATGCT3' where n=3-4 G nucleotides. Library
Source: Michael White, Maria E. Jerome, Emily A. Johnson,
Jay A. Radke, Montana State University. Clone
Availability: David Sibley, Washington University"

ORIGIN
Alignment Scores:
Pred. No.: 545 Length: 545
Score: 190.00 Matches: 64
Percent Similarity: 44.81% Conservative: 31
Best Local Similarity: 30.19% Mismatches: 75
Query Match: 8.40% Indels: 42
DB: 13 Gaps: 9
US-09-985-689A-3 (1-433) x BUS75479 (1-545)
QY 64 AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
DB 4 GAGCACCAGCGCCAGCGCATGTCGCGGC---ATCATCGGGCGCATCAAGACCAA 60
QY 79 AsnAlaLeuAsnGlyGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98
DB 61 AACTCCCGGTGAAGCGGTCTGTGGGAACACAGCATTCGCCGCTTAAGTTCATGGGT 120
QY 99 SerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
DB 121 GCAACCGA-----AACGGGTGCAGTCAAGTCCATCAAGGCTCTCAACTACGCCGTC 174
QY 119 AsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThr 138
DB 175 CAGATGGGAATCCCTCAGCTGCACTGCTGGGGCGGTCCACCTGGTCCGAGCGGTG 234
QY 139 AlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAla 158
DB 235 ATTGCGGCTCTGGAGCGCGGAGAGCGTCGGTCAC-----CTTTTCATTCGCGG 285
QY 159 AlaGlyAsnGluGlyProAsnSerGlyThrIle-----SerAlaProGlyThrAla 175

Db 286 GCGGAAACCAAGGAAGAACACCGAATTCGGCATTACCTCGCTCGTACCGGCTC 345
QY 176 LysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAla 195
Db 346 GCAACAGTAGTCAGCGTGGCTGGACA----- 372
QY 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db 373 AACTCCGAGGACCAACTCGCTCCCTTCAGCAACGGGGAGCGGC----- 417
QY 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 418 ACGGTGACCTTGGCGCGCGCGCGCTTAAATCTCT-----TCTACCTTCGCGCA 468
QY 236 AspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMet 255
Db 469 GAC-----CAGTTTCGGAACCTTCCGGAACCTCCATG 501
QY 256 AlaThrProIleValAlaGlyAsnValAlaGlnLeu 267
Db 502 GCTACGCGCGTGGTGGCGGGCGTCCGCGCATCTCTC 537

RESULT 11
TA315H10P 574 bp DNA linear GSS 13-DEC-2000
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 315h10, forward sequence.
genomic survey sequence.
ACCESSION
AL490202
VERSION
AL490202.1 GI:11866292
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei
ORGANISM
Trypanosoma brucei
Bukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
1 (bases 1 to 574)
AUTHORS
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajadream, M.A. and Barrell, B.G.
TITLE
Direct SubMISSION
JOURNAL
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: released@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/

FEATURES
source
1. 574
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clones="315h10"
ORIGIN
Alignment Scores:
Pred. No.: 9.72e-09 Length: 574
Score: 188.00 Matches: 51
Percent Similarity: 47.65% Conservative: 30
Best Local Similarity: 30.00% Mismatches: 59
Query Match: 8.31% Indels: 30
DB: 29 Gaps: 6